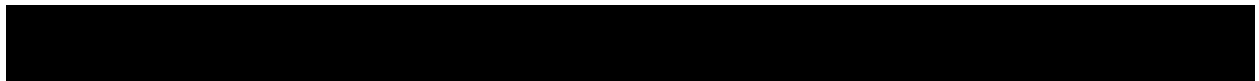


**United States Environmental
Protection Agency**

**Office of Research and
Development
Washington, DC 20460**

**EPA 600/R-01/048
August 2001**

SMOKE Tool for Models-3 Version 4.1 Structure and Operation Documentation



SMOKE Tool for Models-3 Version 4.1 Structure and Operation Documentation

Atmospheric Modeling Division *
National Exposure Research Laboratory
Research Triangle Park, NC 27711

and

EPA Systems Development Center
(A Contractor Operated Facility)
Science Applications International Corporation
6565 Arlington Boulevard, Suite 300
Falls Church, VA

Contract No. 68-W-99-002, Task Order 10

Task Order Project Officer

William G. Benjey
Atmospheric Modeling Division*

*On Assignment from the National Oceanic and Atmospheric Administration,
U.S. Department of Commerce

National Exposure Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

NOTICE

This publication is a part of the documentation the Models-3 Third Generation Air Quality Modeling System. The U.S. Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the procedures described here under contract 68-W-99-002 to Science Applications International Corporation. This document has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products is not intended to constitute endorsement or recommendation for use. The SMOKE Tool was developed for release with Version 4.1 of the Models-3 air quality modeling system under contract to the U.S. Environmental Protection Agency and is free of restrictions to users. Users are requested to provide copies of derivative works to the government without restrictions as to use by others. Users are responsible for acquiring their own copies of commercial software associated with the SMOKE Tool portion of Models-3 release and are also responsible to those vendors for complying with any of the vendors' copyright and licence restrictions. Copies and documentation of Models-3 obtained from the National Technical Information Service (NTIS) are subject to NTIS cost-recovery charges. Portions of the I/O API and SMOKE used with SMOKE Tool, and the Models-3 model builder are Copyrighted 1993-2001 by MCNC- North Carolina Supercomputing Center, and are used with their permissions subject to the above restrictions. Rights and restrictions for copyrights and trademarks used throughout this document are listed in *Appendix E* of the Models-3 Version 4.1 Installation Manual. Although efforts were made to make Models-3 user friendly, it is a complex software system. Consequently, it is important that the system's administrator read the Models-3 Installation Manual before installing the system, and this document before using SMOKE Tool. Installation instructions must be followed closely. There are error conditions which can come from third-party software, and are not Models-3 problems specifically. For example, expired or unavailable SAS or Arc/Info licences or versions not tested with Models-3 can result in error messages.

ABSTRACT

The SMOKE Tool is an emission modeling component of Models-3, a flexible system designed to simplify the development and use of air quality models and other environmental decision support tools. Models-3 is designed for applications ranging from regulatory and policy analysis to understanding the complex interactions of atmospheric chemistry and physics. The summer 2001 release of Models-3 contains a Community Multi-scale Air Quality (CMAQ) modeling system for urban to regional scale air quality simulation of tropospheric ozone, acid deposition, visibility, and fine particles. One of the principal changes to the Models-3 system is the replacement of the former emission modeling system with provision to use the more flexible and efficient Sparse Matrix Operator Kernel Emission (SMOKE) modeling system (available at no cost from MCNC, North Carolina Supercomputing Center), and the SMOKE Tool to create input files for SMOKE.

This document provides details of the SMOKE Tool architecture, system requirements, file formats, environment variables for loading of data sets. The SMOKE can not format and prepare its own emission data input files, which must be gridded and in a particular format. The SMOKE Tool, which is based on the SAS® programming language, imports and performs quality control on emission data related files, allows analysis of the data using the Arc/Info® geographical information system (GIS), places emission inventory data files in the IDA (independent data analysis) format consistent with the Net CDF I/O API data format used by Models-3, and grids the data in accordance with a user defined grid definition from Models-3.

FOREWORD

The Models-3 Community Multi-scale Air Quality Model (CMAQ) modeling system has been developed and improved under the leadership of the Atmospheric Modeling Division of the EPA National Exposure Research Laboratory in Research Triangle Park, NC. The SMOKE Tool along with the SMOKE emission modeling system, is an important part of the summer 2001 Sun UNIX and NT-based releases (Version 4.1) of Models-3. It represents significant improvements to the emission modeling portion of the system. Other than third-party costs as described in the Notice, the SMOKE Tool portion of Models-3 is available without charge for use by air-quality regulators, policy makers, industry, and scientists to address multi-scale, multi-pollutant air quality concerns.

In keeping with its capabilities Models-3, including the SMOKE Tool, is a sophisticated and complex system.. Consequently, before using the SMOKE Tool, the system administrator is advised to read the Models-3 Version 4.1 System Installation and Operation Manual, as well as this documentation. New components of the Models-3 system will improve the capability and flexibility for multi-pollutant air quality modeling. The SMOKE Tool, in concert with SMOKE, represents another technical aid in the process of improving Models-3 capabilities in and allowing a truly multi-pollutant approach to air quality studies.

William B. Petersen
August 2001

CONTENTS

NOTICE	ii
ABSTRACT	ii
FOREWORD	iii
1.0 SMOKE TOOL	1
1.1 Emission Chemical Mechanism	1
1.1.1 Emission Mechanism Outputs	2
1.1.2 Speciation Profile Table Inputs	3
1.1.2.1 Carbon-4-Type Mechanism Lumping Inputs	4
1.1.2.2 RADM Type Mechanism Lumping Inputs	5
1.2 Grid Spatial Surrogate Preparation	6
1.2.1 Grid Directories	9
1.2.2 Coverage Definition Table	9
1.2.3 GIS Coverages	10
1.2.4 Adding Coverages to Spatial Surrogate Processor	12
1.2.5 Gridded Data Files	13
1.2.6 Feature-Surrogate Table	14
1.2.7 Census Data File	16
1.2.8 Census State-County Fraction Table	18
1.2.9 Spatial Surrogate File	18
1.3 Emission Inventory Preparation	19
1.3.1 NET Format Area Source Inventory	23
1.3.2 IDA Format Area Source Inventory	25
1.3.3 NET Format Point Source Inventory	27
1.3.4 IDA Format Point Source Inventory	30
1.3.5 Motor Vehicle Activity Inventory	33
1.3.6 Motor Vehicle Emission Inventory	34
1.4 SMOKE Input File Generation	36
1.4.1 Temporal Profile Files	36
1.4.2 Temporal Cross-Reference Files	38
1.4.3 Gridding Cross-Reference Files	42
1.4.4 Area and Point Source Control Files	42
1.4.5 Speciation Cross-Reference File	46
1.4.6 Pollutant Definition Files	46
1.4.7 MOBILE Emission Factor Cross-Reference Files	47
1.4.8 EGAS Files	48
1.5 Process for Registering Datasets from within the SMOKE Tool	49

2.0	STUDY PLANNER	49
2.1	Grid Information	50
2.2	Case Information	53
2.3	Chemical Mechanism Information	53
2.4	Extracting CEM Data for Case	59
2.5	File Lists for SMOKE Input	60
APPENDIX A		61
	Running SMOKE Tool Outside of Models-3	61
	Smoketool grid instructions	62
	Define Grid	63
	Generate Coverages	63
	Compute Surrogates	64
	Test Surrogate File	65
	Example Grid (m3grid_c_cmaqmet_36eval)	66

1.0 SMOKE TOOL

The SMOKE Tool is an input processor for SMOKE, executed in part via Strategy Manager when running with the Models-3 framework (see Appendix A for stand alone operation of SMOKE Tool). The tool has four basic functions:

- Emission Chemical Mechanism Preparation.
- Emission Inventory Preparation.
- SMOKE Input File Preparation.
- Grid Spatial Surrogate Preparation.

The main SMOKE Tool screen starts the first three functions listed above. These functions use the Models-3 framework Graphical User Interface (GUI). Individual functions within the SMOKE Tool have the ability to register the datasets produced (in the Models-3 Dataset Manager). The Grid Spatial Surrogate Preparation function is executed in batch mode through the Models-3 Study Planner rather than via Strategy Manager. An overview of the procedures followed in running the SMOKE Tool using the Models-3 framework can be found in: Benje, W.G., M.R. Houyoux, and J.W. Susick, "Implementation of the SMOKE Emission Data Processor and SMOKE Tool Input Data Processor in Models-3", presented at the Tenth Emission Inventory Conference in Denver, CO, May 2, 2001, and available via the EPA Models-3 web site (<http://www.epa.gov/asmdnerl/models3>).

1.1 Emission Chemical Mechanism

The following sections describe the Emission chemical mechanism implemented in the SMOKE Tool. The Emission chemical mechanism supplies speciation profile tables to the emission models. These tables contain profiles that tell emission programs how to split base pollutants into the Emission chemical species of the chemical mechanism. Profiles may either be input directly to the chemical mechanism or computed from a lumping mechanism.

The Emission chemical mechanism provides functions that:

- Enter and update input data.
- Compute or import split factors.
- Pass the mechanism data to emission systems.

When defining a chemical mechanism, users supply a mechanism name and type. The name identifies the mechanism within Models-3 and is used whenever referencing the Emission chemical mechanism in Models-3. The chemical mechanism used by the CMAQ Chemistry Transport Model (CCTM) is linked to an emission mechanism by entering this name when defining the chemical mechanism in the Science Manager.

The mechanism type defines the lumping method (or algorithm) used to compute the split factor profiles. Each lumping method has different input requirements. The following types are supported:

- CB4 Computes split factor profiles using the CB4 lumping method.
- RADM2 Computes split factor profiles using the RADM2 lumping method.
- NONE Directly imports all split factor profiles. Performs no lumping calculations.

The following sections describe the output files produced and the input files required for each mechanism type. Unless otherwise specified, all files are free-format, space-delimited American Standard Code for Information Interchange (ASCII) files. The SMOKE Input File Generation function generates a cross-reference file that maps sources to speciation profiles for use in the SMOKE speciation.

1.1.1 Emission Mechanism Outputs

The Speciation Profile Table is a fixed format ASCII file that defines split factor profiles for the chemical mechanism. The fields on each record are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-5	INPRF	Char	Profile Number
7-22	POLID	Char	Pollutant ID
24-39	SPECID	Char	Species ID
41-53	FACTOR	Real	Split Factor
55-67	DIVISOR	Real	Divisor
69-77	XMF	Real	Mass Fraction

Emissions in moles and grams can be computed from the split factors using the following equations:

$$\begin{aligned}\text{moles} &= \text{emis} * \text{factor} / \text{divisor} \\ \text{grams} &= \text{emis} * \text{xf}\end{aligned}$$

where emis is the base pollutant emission value in grams.

1.1.2 Speciation Profile Table Inputs

The factors in the Speciation Profile Table can be computed using a lumping mechanism or imported directly. The files needed to compute the profiles are:

- Base Profile Table.
- Chemical Data.
- Non-Computed Split Factors.
- Mechanism Type-Specific.

Non-Computed Split Factors is the only file needed when directly importing all profiles (mechanism type NONE). All of the files described in this section are used by the SMOKE Tool to create the Speciation Profile Table and are not read directly by SMOKE.

The Base Profile Table defines the SAROAD chemical compound codes in each Speciation profile. The mass fractions of the SAROAD codes in a profile should total 1. The fields on each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	INPRF	Char	Profile number
2	POLID	Int	Pollutant ID
3	ISRD	Int	SAROAD code
4	XMF	Real	Mass fraction of profile assigned to SAROAD

The Chemical Data file contains a master list of the base chemical compounds that can be used in the chemical mechanism. The fields on each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	ISRD	Int	SAROAD code
2	MLWT	Real	Molecular weight (g/mole)
3	SRDNAM	Char	SAROAD name (contains all characters from field start to end of record)

The Non-Computed Split Factors file defines additional split factors for the chemical mechanism. This file defines split factors that are fed directly into the mechanism and not computed in the

lumping. This file is appended to the Speciation Profile Table for SMOKE to read. The fields on each record are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-5	INPRF	Char	Profile number
7-22	POLID	Char	Pollutant ID
24-39	SPECID	Char	Species ID
41-53	FACTOR	Real	Split Factor (normally mass fraction)
55-67	DIVISOR	Real	Divisor (normally molecular weight)
69-77	XMF	Real	Mass Fraction

Mechanism Type-Specific files are specific to the type of lumping used in the mechanism and are defined in the following sections.

1.1.2.1 Carbon-4-Type Mechanism Lumping Inputs

This section describes the lumping files that are needed specifically for a Carbon-4-type chemical mechanism. This type of mechanism performs chemical lumping using an algorithm based on carbon bonding. The additional input files needed are:

- Model Species Data.
- Lumping Data.

The Model Species Data file defines the chemical species to be generated. The fields on each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	MDLSPCD	Int	CB-IV Species Number (1,2,3,...,n)
2	SPECID	Char	Species ID
3	POLID	Char	Pollutant ID
4	SPECNAM	Char	Species Name (optional, contains all characters from field start to end of record)

The Lumping Data file consists of a header record, followed by one or more lumping data records. (Note: literal variable names are left blank in this documentation.) The format of the header record is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1		Char	'NSPEC'
2	n	Int	Number of CB_IV species

The fields on each lumping data record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	ISRD	Int	SAROAD Code
2	ISRDNC	Int	Number of carbon atoms
3	XNUM1	Real	# of moles of species 1 in one mole of SAROAD
4	XNUM2	Real	# of moles of species 2 in one mole of SAROAD
5	XNUM3	Real	# of moles of species 3 in one mole of SAROAD
.			
.			
.			
n+2	XNUMn	Real	# of moles of species n in one mole of SAROAD
n+3	NAME	Char	SAROAD Name (optional, contains all characters from field start to end of record)

1.1.2.2 RADM Type Mechanism Lumping Inputs

This section describes the Lumping files that are needed specifically for a Regional Acid Deposition Model (RADM)-type chemical mechanism. This type mechanism performs the chemical lumping using an algorithm based upon chemical reactivity. The additional input files needed are:

- SAROAD-Category Table.
- Category-Species Table.

The SAROAD-Category Table maps each SAROAD to a chemical category. The fields on each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	ISRD	Int	SAROAD code
2	CATEGORY	Int	Chemical category number
3	SRDNAM	Char	SAROAD name (optional, contains all characters from field start to end of record)

The Category-Species Table maps categories to chemical species. Allocation factors for the species in a category should total 1. The fields on each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	CATEGORY	Int	Chemical category number
2	SPECID	Char	Species ID
3	ALOC_FAC	Real	Allocation Factor
4	REAC_FAC	Real	Reactivity Factor
5	SPECNAM	Char	Species Name (optional, contains all characters from field start to end of record)

1.2 Grid Spatial Surrogate Preparation

The primary purpose of the grid processor is to generate spatial surrogates for SMOKE horizontal gridding. The grid processor is executed through nodes in a study plan. The underlying functions that the grid processor executes are modified versions of the Statistical Analysis System (SAS) and ArcInfo gridding programs copied from the MEPPS Emission Processor (EMPRO). The grid processor performs the following functions:

- **Define Grid** Creates the grid workspace, receiving the grid description file from the Study Planner. The program creates the grid directories, imports the grid description, and generates the base Geographic Information System (GIS) coverages in the workspace. The grid description is read from file G_GRIDPATH (supplied from the study planner).

- **Generate Coverages** Generates GIS coverages in the workspace. The user selects the coverages to generate through environment variables. The names of the environment variables are based upon the GIS coverages defined to the SMOKE Tool.
- **Generate Surrogates** Computes spatial surrogates for the grid from the GIS coverages generated for the grid. The user selects the coverages to include in the surrogate calculation through environment variables. The program computes all spatial surrogates defined for each coverage selected.

The grid processor creates the grid directories in a user-specified location. The location is specified through the following environment variables:

- **EMS_HOME** Workspace Path. The base directory under which the grid information is stored. Supplied by user.
- **HGRIDNAME** Models-3 Grid Name. The grid name is used to define subdirectories under the workspace path for the grid. Supplied by Study Planner.

All data for the grid, created by the Define Grid and Generate Coverages functions, is stored in subdirectories under directory **\$EMS_HOME/gridspec/\$HGRIDNAME**. The user specifies the location of the spatial surrogate file, created by the Generate Surrogate function, in a plan output file specification.

The user defines the GIS coverages available to SMOKE Tool through the following environment variables:

- **GISDB** Name of the GIS data base directory. This directory contains the GIS coverages and related data files for the surrogate processing.
- **COVER_DEF** Name of the file containing the coverage definition table. This table defines the GIS coverages available for surrogate processing.

The initial GIS coverages released with Models-3 are:

- **COUNTY** County areas
- **CENSUS** Census tract/block group areas (US only)
- **FHAROAD** FHA road lengths (US only)

- AGRICULTURE Agricultural areas (US only)
- AIRPORTS Airport locations
- PORTS Port locations
- RAILROADS Railroad lines
- LAND_WATER Land-water areas
- URBAN-RURAL Urban-rural areas
- ROADS Road lengths
- FOREST Forest areas
- TIGER TIGER/Line roads (US only)

To select a coverage, set the environment variable for the coverage to YES or Y. The value may contain upper or lower case letters. Sample environment variable settings that cause coverages to be processed are:

```

PORTS=Y
AIRPORTS=YES
RAILROADS=yes

```

The surrogates step also uses the following environment variables:

- FEATURE_SRG Name of the Feature-Surrogate Table. This table defines the manner in which spatial surrogates are computed from the coverage features.
- CENSUS_DATA Name of the Census Data File. This file contains housing and population data associated with the CENSUS coverage. It may also contain state or county level data.
- FEAT_STCY_FRAC Name of the Census State-County Fraction table. This file contains fractions to allocate state level census data to the counties within the states.
- SURROGATES Name of the spatial surrogates file. This contains the spatial surrogates (or gridding coefficients) computed in the surrogates step.

The follow sections describe the various directories, GIS coverages, and files used in the spatial surrogate processor.

1.2.1 Grid Directories

The grid master directory is **\$EMS_HOME/gridspec/\$EMS_GRID**. The subdirectories created are shown in the following table:

<u>Subdirectory</u>	<u>Type of Data</u>	<u>Environment Variable</u>	<u>SAS Reference</u>
common/sas	Ungridded Emissions	EMS EMS_CVRT	library EMS file EMS_CVRT
common/gis	Ungridded GIS	EMSG	file EMSG
sas	Gridded Emissions	EMS_GRD EMSF_GRD	library EMS_GRD file EMSF_GRD
gis	Gridded GIS	EMSG_GRD	file EMSG_GRD

The first two directories contain data for the grid's geographic area that have not been processed for a grid. The EMSG directory contains GIS coverages that have not been overlaid with a grid.

The gridded data directories contain data that have been processed for the grid. EMS_GRD contains the grid description and gridded data files. The EMSG_GRD directory contains GIS coverages that have been overlaid with the grid.

1.2.2 Coverage Definition Table

The coverage definition table defines the GIS coverages available to the SMOKE Tool. The table is a free-format, space-delimited ASCII file containing the following information on each record:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COVNAME	Char	SMOKE Tool coverage name. This name is the identifier used for the coverage in the SMOKE Tool. It provides the name of the environment variable used to select the coverage in processing, and is the name used to reference the coverage in SMOKE Tool processing.
2	COUNTRY	Char	Country name for GIS coverage.
3	COVTYPE	Int	Coverage type (poly, line, point).
4	COVERAGE	Real	Coverage name in the GISDB directory. This name defines the location of the coverage to ARC.

A SMOKE Tool coverage can contain one or more GIS coverages. In the simplest instance, the Tool coverage contains a single GIS coverage that contains all countries. In that case, the coverage definition table contains a single record with a country name of “ALL”. The more common instance is that there are individual GIS coverages for each country. In that case, the table will contain an entry for each country. The entries will each contain a country name (US, CANADA, MEXICO). The SMOKE Tool coverage generation treats the set of GIS coverages under a coverage name as a single coverage.

A sample table is shown below:

COUNTY	ALL	poly	na_geo
CENSUS	US	poly	us_bg90
FHAROAD	US	line	fhards_48
AGRICULTURE	US	poly	us-ag
AIRPORTS	US	point	us-ap
AIRPORTS	CANADA	point	can-ap
PORTS	US	point	us-port
PORTS	CANADA	point	can-port
RAILROADS	US	line	us-rr
RAILROADS	CANADA	line	can-rr
LAND_WATER	US	poly	us-lwcty
LAND_WATER	CANADA	poly	can-lwcty
URBAN_RURAL	US	poly	us-urcty
URBAN_RURAL	CANADA	poly	can-urcty
ROADS	US	line	us-rd
ROADS	CANADA	line	can-rd
FOREST	US	poly	us-for
FOREST	CANADA	poly	can-for

1.2.3 GIS Coverages

The GISDB directory contains the GIS coverages used by the SMOKE Tool. The GISDB directory contains two types of coverages:

- Coverages from the original SMOKE Tool spatial surrogate processor. These include the COUNTY and FHAROAD coverages. TIGER road coverages are also included, but are located in a separate directory (TIGERDB).
- Universal coverages. All user provided coverages must be universal coverages. A universal coverage is a user-provided GIS coverage not pre-defined in the SMOKE Tool source code. Individual coverages can define features for one country or a set of countries. The coverages do not have to contain the entire country, but must contain the section of the country needed for SMOKE Tool and SMOKE processing. The SMOKE Tool coverage definition file ties country coverages of a specific type together for processing. There can be only one entry per coverage name and country combination. Consequently, if two coverages of the same type (including the same kinds of information)

for different sections of a country were to be used, they must be merged into a single GIS coverage for use in the SMOKE Tool.

Universal coverages must have the following projection parameters:

- Projection Geographic
- Units DD
- Datum NAD83
- Spheroid GRS1980

and contain the following data items:

- STATE Country/state code. For US coverages this is the 2-digit FIPS state code. For other countries, this is a 3-digit field with the country code in the first digit and the state (or province) code in the second and third codes.
- COUNTY County code. This is a 3-digit field for all coverages. The field contains the FIPS county code for US coverages.
- STIDCYID Combined country/state/county code. This field may be a redefinition of the STATE and COUNTY fields. For US coverages this may be a 5-digit field. For other countries it must be a 6-digit integer field.
- COV_CODE Feature code. Classifies features within the coverage. May be zero if the coverage only has one feature. In the CENSUS coverage, this field contains the censusid. For US coverages this will either be a census tract or block group ID. The feature codes for the CENSUS coverage are defined in the CENSUS_DATA file included with the coverage, which contains the housing and population data associated with the coverage. The field length can be one to sixteen digits, based upon the coverage.

The feature codes defined in the original universal coverages are shown below:

<u>Coverage</u>	<u>Code</u>	<u>Description</u>
AGRICULTURE	1	Agriculture Area
	2	Non-Agriculture Area
AIRPORTS	0	Airport Location
PORTS	0	Port Location
RAILROADS	0	Railroad

<u>Coverage</u>	<u>Code</u>	<u>Description</u>
LAND_WATER	1	Water Area
	2	Island Area
	3	Mainland Area
URBAN_RURAL	1	Urban Area
	2	Rural Area
ROADS	01	Unclassified Urban Road
	02	Unclassified Rural Road
	11	Urban Primary Road
	12	Rural Primary Road
	21	Urban Secondary Road
	22	Rural Secondary Road
FOREST	1	Forest Area
	2	Non-forest Area

1.2.4 Adding Coverages to Spatial Surrogate Processor

The basic steps for adding a coverage to the SMOKE Tool spatial surrogate processor are:

- Prepare the coverage (in Arc Info prior to using SMOKE Tool).
- Copy the new coverage into the GISDB directory using ArcInfo.
- Update the coverage definition table to include the new coverage.
- Update the feature-surrogate table to include surrogates from the new coverage

When preparing the coverage the user will need to perform the following actions:

- Determine which fields already present in the coverage will be used to populate the stidcyid, state, county, and cov_code required fields. These fields must be type Integer, and match the codes in the SMOKE Tool county table.
- Populate the required fields with these values.

- Project the coverage to the following parameters:

–	Projection	Geographic
–	Units	DD
–	Datum	NAD83
–	Spheroid	GRS1980

After preparing the coverage, the user must copy the coverage into the GISDB directory. The Arc command 'copy' must be used to copy the coverage. The command syntax is:

```
copy cover new-cover
```

where *new-cover* includes the path and name of the coverage. The path will be the GISDB directory.

The coverage definition table defines the GIS coverage to the SMOKE Tool. A line in the table will have to be changed or added to include the new coverage. The example below adds a Mexican AGRICULTURE coverage (*mex_ag*). The table currently has US and Canadian AGRICULTURE coverages. In this case, the user will add the following line:

<u>Covname</u>	<u>Country</u>	<u>Type</u>	<u>Coverage</u>
AGRICULTURE	MEXICO	poly	mex_ag

The feature codes (*cov_code*) of a coverage, being added to an existing SMOKE Tool coverage set, must match the feature codes of the other coverages in the set.

The final step is to define the spatial surrogates computed by the coverage. This is done by adding entries in the feature-surrogate table for the coverage set. Since the previous example adds a country to an existing coverage set, the spatial surrogates are already defined. When a new coverage set is added, entries must be added that define the spatial surrogates to be computed from the coverage features.

1.2.5 Gridded Data Files

The coverages step generates gridded data files in the EMS_GRD directory for each processed. It generates two files for each universal coverage:

- *covname.cy* County level gridded data
- *covname.st* State level gridded data

The gridded data files are free-format, comma-delimited ASCII files. The fields on each file record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	STID	Int	Country/StateCode
2	CYID	Int	County code (zero for state level data)
3	FEATURE	Char	Feature Code
4	COL	Int	Grid column number (zero if out of grid)
5	ROW	Int	Grid row number (zero if out of grid)
6	GVAL	Real	Size of feature in county (or state)-grid cell intersection
7	CYVAL	Real	Size of feature in county (or state)

The coverages step only generates one gridded data file (census.cy) for the CENSUS coverage. It is also a free-format, comma-delimited ASCII file. In place of the feature code, the file contains the censusid, which in the US coverage is a census block group ID.

1.2.6 Feature-Surrogate Table

The Feature-Surrogate table defines the manner in which spatial surrogates are computed from the coverage features. The table is a free-format, space-delimited ASCII file containing the following information on each record:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COVNAME	Char	SMOKE Tool coverage name. This name is the identifier used for the coverage in the SMOKE Tool. It provides the name of the environment variable used to select the coverage in processing, and is the name used to reference the coverage in SMOKE Tool processing
2	FEATURE	Char	Coverage feature code
3	SSC	Int	Spatial surrogate code
4	FACTOR	Real	Fraction of feature added to spatial surrogate (default is one)

A feature may be included with more than one spatial surrogate, either entirely or partially. A sample table is shown below:

#	Coverage	Feature Code	to	Spatial	Surrogate Code Table
#	covname	feature	ssc	factor	
	AGRICULTURE	1	1	1	Agriculture (US)
	CENSUS	AGAREA	1	1	Agriculture (Canada)
	AIRPORTS	0	2	1	Airports
	LAND_WATER	2	3	1	Land Area (Island)
	LAND_WATER	3	3	1	(Mainland)
	CENSUS	TOTHU	4	1	Total Housing Units
	ROADS	11	7	1	Major Highways (Urban Primary Roads)
	ROADS	12	7	1	(Rural Primary Roads)
	CENSUS	TOTPOP	8	1	Total Population
	PORTS	0	9	1	Ports
	RAILROADS	0	10	1	Railroads
	LAND_WATER	1	11	1	Water Area
	URBAN_RURAL	2	12	1	Rural Area
	URBAN_RURAL	1	13	1	Urban Area
	FOREST	1	14	1	Forest Area
	ROADS	11	15	1	Urban Primary Roads
	ROADS	12	16	1	Rural Primary Roads
	ROADS	21	17	1	Urban Secondary Roads
	ROADS	22	18	1	Rural Secondary Roads
	CENSUS	URBPOP	19	1	Urban Population
	CENSUS	RURPOP	20	1	Rural Population

Except for the CENSUS coverage, each table entry refers to a feature code defined in the coverage, and returned on the gridded data file generated for the coverage. The spatial surrogates are computed using the gridded areas, lengths, or point counts returned for the coverage features.

The CENSUS entries refer to feature codes defined on the census data file. CENSUS surrogates are computed by applying census data to gridded areas returned from the CENSUS and COUNTY coverages. Censused level data is applied to the gridded CENSUS data. County level data is applied to the gridded COUNTY areas. State level data is first apportioned to counties and then applied to the gridded COUNTY areas.

The sample table shown above generates the following spatial surrogates:

<u>SSC</u>	<u>Description</u>
1	Agriculture
2	Airports
3	Land Area
4	Total Housing Units
7	Major Highways
8	Total Population

<u>SSC</u>	<u>Description</u>
9	Ports
10	Railroads
11	Water Area
12	Rural Area
13	Urban Area
14	Forest Area
15	Urban Primary Roads
16	Rural Primary Roads
17	Urban Secondary Roads
18	Rural Secondary Roads
19	Urban Population
20	Rural Population

1.2.7 Census Data File

The census data file contains data that can be used in conjunction with spatial data at the censusid or county level to compute spatial surrogates. The file is a free-format, space-delimited ASCII file with the following information on each record:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	STID	Int	Country/State/County Code
2	CYID	Int	County code (zero indicates missing)
3	CENSUSID	Char	Census ID (zero indicates missing)
4	FEATURE	Char	Feature Code
5	VALUE	Real	Feature Value

The data can be entered at the state, county, or censusid level. Censusid data has a stid, cyid, and censusid. County level data has the stid and cyid, but no censusid. State level data only has the stid.

The feature codes in the current census data files are:

- TOTHU Total housing units
- TOTPOP Total population
- RURPOP Rural population
- URBPOP Urban population
- AGAREA Agricultural area (used for Canada)

Sample US and Canadian data is shown below. The US data is at the census block group level. The Canadian data is at the province level.

```

1 1 02011  RURPOP  65
1 1 02011  TOTHU   370
1 1 02011  TOTPOP 1105
1 1 02011  URBPOP 1040
1 1 02012  RURPOP  50
1 1 02012  TOTHU   301
1 1 02012  TOTPOP  756
1 1 02012  URBPOP  706
1 1 02021  RURPOP   0
1 1 02021  TOTHU   273
1 1 02021  TOTPOP  783
1 1 02021  URBPOP  783
1 1 02022  RURPOP   0
1 1 02022  TOTHU   419
1 1 02022  TOTPOP 1194
1 1 02022  URBPOP 1194
.
.
.
160 0 0  RURPOP 32085
160 0 0  TOTHU 11813
160 0 0  TOTPOP 32085
160 0 0  URBPOP  0
161 0 0  RURPOP 1218376
161 0 0  TOTHU 306468
161 0 0  TOTPOP 1218376
161 0 0  URBPOP  0
111 0 0  AGAREA 141021.2778
112 0 0  AGAREA 12369.2827
113 0 0  AGAREA 684.8532
124 0 0  AGAREA 27594.2432
135 0 0  AGAREA 90992.9167
146 0 0  AGAREA 103307.607
147 0 0  AGAREA 1214942.9943
148 0 0  AGAREA 10979.4997
159 0 0  AGAREA 62492.9607

```

1.2.8 Census State-County Fraction Table

This table supplies factors that allocate state level census data to counties. The file is a free-form space-delimited ASCII file with the following information on each record:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	FEATURE	Char	Feature Code
3	FRAC	Real	Fraction of state feature value in county

The following lines allocate Newfoundland (province 110) total population to counties for the spatial surrogate calculation:

```
110001 TOTPOP 0.44540757
110002 TOTPOP 0.05161995
110003 TOTPOP 0.04263404
110004 TOTPOP 0.04519323
110005 TOTPOP 0.07971190
110006 TOTPOP 0.07077839
110007 TOTPOP 0.07594072
110008 TOTPOP 0.09126545
110009 TOTPOP 0.04401675
110010 TOTPOP 0.05343200
```

1.2.9 Spatial Surrogate File

The Spatial Surrogate files are free-format, space-delimited ASCII files. Each file consists of a header record followed by one or more detail records. The header record contains a description of the grid. The detail records contains spatial surrogate data for the grid. The file contains spatial records for both states and counties. The format of the header record is:

```
#GRID      param1 param2 ... param15
```

The header record parameters are show below:

<u>Parameter</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	GDNAME	Char	Grid name (up to 16 characters)
2	XORIG	Real	X origin (projection units)
3	YORIG	Real	Y origin (projection units)

<u>Parameter</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
4	XCELL	Real	Cell size, X direction (projection units)
5	YCELL	Real	Cell size, Y direction (projection units)
6	NCOLS	Int	Number of cells in X direction
7	NROWS	Int	Number of cells in Y direction
8	NTHIK	Int	Boundary thickness (number of cells)
9	PROJTYPE	Char	Project type (UTM, LAMBERT, LAT-LON,...)
10	PROJUNIT	Char	Projection units (DEGREES or METERS)
11	P_ALP	Real	First map projection description parameter
12	P_BET	Real	Second map projection description parameter
13	P_GAM	Real	Third map projection description parameter
14	P_XCENT	Real	Center of coordinate system (degrees longitude)
15	P_YCENT	Real	Center of coordinate system (degrees latitude)

The format of the spatial surrogate records is shown below:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	SSC	Int	Spatial Surrogate Code
2	COUNTY	Int	Country/State/County Code
3	COL	Int	Grid Column Number (x-axis)
4	ROW	Int	Grid Row Number (y-axis)
5	RATIO	Real	Spatial Surrogate Ratio (fraction of county value in grid cell)
6	VALUE	Real	Surrogate Value for county-grid cell intersection

1.3 Emission Inventory Preparation

The inventory processor consists of SAS screens and programs that import, quality assure (QA), and export emission inventories. The functions are:

- Create/Delete working directories. Each working directory has a geographic area and data year associated with it.
- Import inventory data. Loads data from a National Emissions Trends (NET) or IDA inventory into the working directory, selecting data for the directory geographic area.
- QA inventory data. Provides error checking and correction, statistical analysis, and data browsing.
- Export inventory data. Creates NET or IDA inventories from the inventory data in a working data directory.
- Define stack splits. Classifies point source stacks as Major Elevated Point Source Emission (MEPSE), major, or minor.

The inventory formats handled by the inventory processor are:

- NET Format Area Source Emission Inventory.
- IDA Format Area Source Emission Inventory.
- NET Format Point Source Emission Inventory.
- IDA Format Point Source Emission Inventory.
- NET Format Motor Vehicle, Vehicle Miles Traveled (VMT) Inventory.
- IDA Format Motor Vehicle VMT Inventory.

The area and point source inventories contain emission data, and the motor vehicle inventories contain VMT data. The formats of these inventories are discussed in later sections of this document. The formats of the area and point source inventories have been modified to be consistent with the 1996 NET format inventories. In the area source inventories, the rule penetration rate field has been changed from three to six characters. In the point source inventories, the boiler ID field has been changed from five to six characters.

The inventory processor does not handle any land use inventories for the biogenics module, but the land use inventory incorporated into MEPPS is available to SMOKE. This is a county-based inventory that contains land use data for the U.S., Canada, and Mexico in the standard Biogenic Emissions Inventory System (BEIS-2) format. The emission factor tables, which are in the standard BEIS-2 format, are also available.

For the export functions, the data from multiple working directories can be written to a single output file. An append function allows data to be added to the end of an existing file. With the append option specified, all records (including headers) are written to the end of an existing file. The motor vehicle export function places header records at the beginning of the inventory data. This is needed if files are combined.

The function that defines stack splits creates the following two fixed-format ASCII files that can be used by SMOKE to define the MEPSE, major, and minor stacks:

- Stack Group Split Definition File.
- Stack Split Definition File.

MEPSE stacks are supplied to CCTM in stack groups; major stacks are supplied at the individual stack level. A stack group is a set of stacks within a plant that have similar stack parameters. All stacks in a group are either MEPSE or non-MEPSE. If a stack is not in a MEPSE group, it can be classified as major or minor.

The Stack Group Split Definition file contains a record for each stack group found in the inventory. The field MEPSE specifies whether the group is MEPSE or not. When a stack is MEPSE, all stacks in the group are MEPSE. The fields in each record are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-6	IGROUP	Int	Stack Group Number
7	MEPSE	Char	MEPSE Switch
8	CTRYID	Int	Country Code
9-10	STID	Int	State Code
11-13	CYID	Int	County Code
14-28	PLANTID	Char	Plant ID
29-33	STKCNT	Int	Number of stacks in group
34-42	LONGITUDE	Real	Stack Longitude (degrees)
43-51	LATITUDE	Real	Stack Latitude (degrees)
55-62	STKDM	Real	Inside stack diameter (meters)
63-69	STKHT	Real	Stack height (meters)
70-76	STKTK	Real	Stack exit temperature (degrees K)
77-83	STKVE	Real	Stack exit velocity (meters/sec)
84-93	STKFLW	Real	Stack exit flow rate (meters**3/sec)

The Stack Split Definition file contains a record for each stack in the working directory inventory. The fields MAJOR and MEPSE specify whether the stack is MEPSE, major, or minor. A

MEPSE stack will always have a major/minor classification that should be ignored when generating output files, since the stack has already been classified as MEPSE. The fields in each record are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-6	ISTACK	Int	Stack Number
7-12	IGROUP	Int	Stack Group Number
13	MAJOR	Char	Major/Minor Switch
14	MEPSE	Char	MEPSE Switch
15	CTRYID	Int	Country Code
16-17	STID	Int	State Code
18-20	CYID	Int	County Code
21-35	PLANTID	Char	Plant ID
36-50	POINTID	Char	Point ID
51-62	STKID	Char	Stack ID
69-77	LONGITUDE	Real	Stack Longitude (degrees)
78-86	LATITUDE	Real	Stack Latitude (degrees)
90-97	STKDM	Real	Inside stack diameter (meters)
98-104	STKHT	Real	Stack height (meters)
105-111	STKTK	Real	Stack exit temperature (degrees K)
112-118	STKVE	Real	Stack exit velocity (meters/sec)
119-128	STKFLW	Real	Stack exit flow rate (meters**3/sec)

The fields MEPSE and MAJOR in the files specify whether stacks are MEPSE, major, or minor. The values for MEPSE are:

- Non-blank - MEPSE.
- Blank - non-MEPSE.

The values for MAJOR are:

- Non-blank - major stack.
- Blank - minor stack.

The Split Definition files do not contain any grid information. The stack and group numbers are arbitrarily assigned from the stacks and groups found in the working directory point source inventory. These files offer two major advantages over simple lists of major stacks and MEPSE stack groups:

- Stack and stack group classifications can be changed by manually editing the files.
- Aggregations of stacks into stack groups are defined.

1.3.1 NET Format Area Source Inventory

The NET format area source inventory is a fixed-format ASCII file. The fields in each record are shown below:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-2	STID	Int	State Code
3-5	CYID	Int	County Code
6-15	SCC	Char	Source Classification Code
16-25	VOC_ANN	Real	VOC Annual Emissions (short tons)
26-35	NOX_ANN	Real	NOX Annual Emissions (short tons)
36-45	CO_ANN	Real	CO Annual Emissions (short tons)
46-55	SO2_ANN	Real	SO ₂ Annual Emissions (short tons)
56-65	PM10_ANN	Real	PM10 Annual Emissions (short tons)
66-75	PM25_ANN	Real	PM2.5 Annual Emissions (short tons)
76-85	NH3_ANN	Real	NH ₃ Annual Emissions (short tons)
86-95	VOC_OSD	Real	VOC Ozone Season Emissions (tons/day)
96-105	NOX_OSD	Real	NOX Ozone Season Emissions (tons/day)
106-115	CO_OSD	Real	CO Ozone Season Emissions (tons/day)
116-125	SO2_OSD	Real	SO ₂ Ozone Season Emissions (tons/day)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
126-135	PM10_OSD	Real	PM10 Ozone Season Emissions (tons/day)
136-145	PM25_OSD	Real	PM2.5 Ozone Season Emissions (tons/day)
146-155	NH3_OSD	Real	NH ₃ Ozone Season Emissions (tons/day)
156-166	VOC_EMF	Real	VOC Emission Factors (SCC units)
167-177	NOX_EMF	Real	NOX Emission Factors (SCC units)
178-188	CO_EMF	Real	CO Emission Factors (SCC units)
189-199	SO2_EMF	Real	SO ₂ Emission Factors (SCC units)
200-210	PM10_EMF	Real	PM10 Emission Factors (SCC units)
211-221	PM25_EMF	Real	PM2.5 Emission Factors (SCC units)
222-232	NH3_EMF	Real	NH ₃ Emission Factors (SCC units)
233-239	VOC_CE	Real	VOC Control Efficiency (%)
240-246	NOX_CE	Real	NOX Control Efficiency (%)
247-253	CO_CE	Real	CO Control Efficiency (%)
254-260	SO2_CE	Real	SO ₂ Control Efficiency (%)
261-267	PM10_CE	Real	PM10 Control Efficiency (%)
268-274	PM25_CE	Real	PM2.5 Control Efficiency (%)
275-281	NH3_CE	Real	NH ₃ Control Efficiency (%)
282-284	VOC_RE	Real	VOC Rule Effectiveness (%)
285-287	NOX_RE	Real	NOX Rule Effectiveness (%)
288-290	CO_RE	Real	CO Rule Effectiveness (%)
291-293	SO2_RE	Real	SO ₂ Rule Effectiveness (%)
294-296	PM10_RE	Real	PM10 Rule Effectiveness (%)
297-299	PM25_RE	Real	PM2.5 Rule Effectiveness (%)
300-302	NH3_RE	Real	NH ₃ Rule Effectiveness (%)
303-308	VOC_RP	Real	VOC Rule Penetration (%)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
309-314	NOX_RP	Real	NOX Rule Penetration (%)
315-320	CO_RP	Real	CO Rule Penetration (%)
321-326	SO2_RP	Real	SO ₂ Rule Penetration (%)
327-332	PM10_RP	Real	PM10 Rule Penetration (%)
333-338	PM25_RP	Real	PM2.5 Rule Penetration (%)
339-344	NH3_RP	Real	NH ₃ Rule Penetration (%)

1.3.2 IDA Format Area Source Inventory

The IDA format area source inventory consists of a set of fixed-format detail records preceded by a set of header records. The general format of the header records is:

```
command parm1 parm2 ... parm n
```

The command field starts in the first position. The remainder of the header is free-format, blank delimited. The specific header formats are:

```
#TYPE filetype
#COUNTRY country-name
#YEAR datayear
#DESC description
#POLID pol1 pol2 ... poln
```

The filetype and description fields are taken as the first non-blank character after the command to the last character of the record. Valid country names are US, CANADA, and MEXICO. Datayear must be a number from 1900 to 2200. The #POLID command defines the pollutants that are contained on the detail records. "pol1" is the ID of the first pollutant. "pol2" is the ID of the second pollutant, etc.

Sample header records are as follows:

```
#TYPE Area Source Emission Inventory
#COUNTRY US
#YEAR 1995
#DESC Alabama 1995
#POLID CO NH3 NOX PM10 PM2_5 SO2 VOC
```

The detail records are fixed-format. The fields in each record are as follows:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-2	STID	Int	State Code
3-5	CYID	Int	County Code
6-15	SCC	Char	Source Classification Code
16-25	ANN1	Real	pol1 Annual Emissions (short tons)
26-35	OSD1	Real	pol1 Ozone Season Emissions (tons/day)
36-46	EMF1	Real	pol1 Emission Factors (SCC units)
47-53	CE1	Real	pol1 Control Efficiency (%)
54-56	RE1	Real	pol1 Rule Effectiveness (%)
57-62	RP1	Real	pol1 Rule Penetration (%)
63-72	ANN2	Real	pol2 Annual Emissions (short tons)
73-82	OSD2	Real	pol2 Ozone Season Emissions (tons/day)
83-93	EMF2	Real	pol2 Emission Factors (SCC units)
94-100	CE2	Real	pol2 Control Efficiency (%)
101-103	RE2	Real	pol2 Rule Effectiveness (%)
104-109	RP2	Real	pol2 Rule Penetration (%)
110-120	ANN3	Real	pol3 Annual Emissions (short tons)
121-130	OSD3	Real	pol3 Ozone Season Emissions (tons/day)
131-141	EMF3	Real	pol3 Emission Factors (SCC units)
142-147	CE3	Real	pol3 Control Efficiency (%)
148-150	RE3	Real	pol3 Rule Effectiveness (%)
151-156	RP3	Real	pol3 Rule Penetration (%)

(Repeat for n pollutants.)

1.3.3 NET Format Point Source Inventory

The NET Format Point Source Inventory is a fixed-format ASCII file. The fields in each record are shown below:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-2	STID	Int	State Code
3-5	CYID	Int	County Code
6-20	PLANTID	Char	Plant Identification Code
21-35	POINTID	Char	Point Identification Code
36-47	STACKID	Char	Stack Identification Code
48-53	ORISID	Char	DOE Plant ID
54-59	BLRID	Char	Boiler Identification Code
60-61	SEGMENT	Char	DOE ID
62-101	PLANT	Char	Plant Name
102-111	SCC	Char	Source Classification Code
112-115	STKHGT	Real	Stack Height (ft)
116-121	STKDIAM	Real	Stack Diameter (ft)
122-126	STKTEMP	Real	Stack Gas Exit Temperature (deg F)
127-135	STKFLOW	Real	Stack Gas Flow Rate (ft ³ /sec)
136-144	STKVEL	Real	Stack Gas Exit Velocity (ft/sec)
145-152	BOILCAP	Real	Design Capacity (mmBtu/hr)
153	CAPUNITS	Char	Capacity Unit Code
154-155	WINTHRU	Real	Winter throughput (% of Annual)
156-157	SPRTHRU	Real	Spring throughput (% of Annual)
158-159	SUMTHRU	Real	Summer throughput (% of Annual)
160-161	FALTHRU	Real	Fall throughput (% of Annual)
162-163	HOURS	Int	Normal Operating Time (hours/day)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
164-165	START	Int	Normal Operation Start Time
166	DAYS	Int	Normal Operating Time (days/week)
167-168	WEEKS	Int	Normal Operating Time (weeks/year)
169-179	THRUPUT	Real	Throughput Rate (SCC units/yr)
180-191	MAXRATE	Real	Max O3 Season Rate (units/day)
192-199	HEATCON	Real	Heat Content (mmBtu/SCC unit)
200-204	SULFCON	Real	Sulfur Content (mass percent)
205-209	ASHCON	Real	Ash Content (mass percent)
210-218	NETDC	Real	Max Nameplate Capacity (MW)
219-222	SIC	Int	Standard Industrial Classification Code
223-231	LATC	Real	Latitude (decimal degrees)
232-240	LONC	Real	Longitude (decimal degrees)
241-250	VOC_EMF	Real	VOC Emission Factors (SCC units)
251-260	NOX_EMF	Real	NOX Emission Factors (SCC units)
261-270	CO_EMF	Real	CO Emission Factors (SCC units)
271-280	SO2_EMF	Real	SO ₂ Emission Factors (SCC units)
281-290	PM10_EMF	Real	PM10 Emission Factors (SCC units)
291-300	PM25_EMF	Real	PM2.5 Emission Factors (SCC units)
301-310	NH3_EMF	Real	NH ₃ Emission Factors (SCC units)
311-317	VOC_CE	Real	VOC Control Efficiency (%)
318-324	NOX_CE	Real	NOX Control Efficiency (%)
325-331	CO_CE	Real	CO Control Efficiency (%)
332-338	SO2_CE	Real	SO ₂ Control Efficiency (%)
339-345	PM10_CE	Real	PM10 Control Efficiency (%)
346-352	PM25_CE	Real	PM2.5 Control Efficiency (%)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
353-359	NH3_CE	Real	NH ₃ Control Efficiency (%)
360-362	VOC_CPRI	Real	VOC Primary Control Equipment Code
363-365	NOX_CPRI	Real	NOX Primary Control Equipment Code
366-368	CO_CPRI	Real	CO Primary Control Equipment Code
369-371	SO2_CPRI	Real	SO ₂ Primary Control Equipment Code
372-374	PM10_CPRI	Real	PM10 Primary Control Equipment Code
375-377	PM25_CPRI	Real	PM2.5 Primary Control Equipment Code
378-380	NH3_CPRI	Real	NH ₃ Primary Control Equipment Code
381-383	VOC_CSEC	Real	VOC Secondary Control Equipment Code
384-386	NOX_CSEC	Real	NOX Secondary Control Equipment Code
387-389	CO_CSEC	Real	CO Secondary Control Equipment Code
390-392	SO2_CSEC	Real	SO ₂ Secondary Control Equipment Code
393-395	PM10_CSEC	Real	PM10 Secondary Control Equipment Code
396-398	PM25_CSEC	Real	PM2.5 Secondary Control Equipment Code
399-401	NH3_CSEC	Real	NH ₃ Secondary Control Equipment Code
402-414	VOC_ANN	Real	VOC Annual Emissions (short tons)
415-427	NOX_ANN	Real	NOX Annual Emissions (short tons)
428-440	CO_ANN	Real	CO Annual Emissions (short tons)
441-453	SO2_ANN	Real	SO ₂ Annual Emissions (short tons)
454-466	PM10_ANN	Real	PM10 Annual Emissions (short tons)
467-479	PM25_ANN	Real	PM2.5 Annual Emissions (short tons)
480-492	NH3_ANN	Real	NH ₃ Annual Emissions (short tons)
493-505	VOC_OSD	Real	VOC Ozone Season Emissions (tons/day)
506-518	NOX_OSD	Real	NOX Ozone Season Emissions (tons/day)
519-531	CO_OSD	Real	CO Ozone Season Emissions (tons/day)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
532-544	SO2_OSD	Real	SO ₂ Ozone Season Emissions (tons/day)
545-557	PM10_OSD	Real	PM10 Ozone Season Emissions (tons/day)
558-570	PM25_OSD	Real	PM2.5 Ozone Season Emissions (tons/day)
571-583	NH3_OSD	Real	NH ₃ Ozone Season Emissions (tons/day)
584-586	VOC_RE	Real	VOC Rule Effectiveness (%)
587-589	NOX_RE	Real	NOX Rule Effectiveness (%)
590-592	CO_RE	Real	CO Rule Effectiveness (%)
593-595	SO2_RE	Real	SO ₂ Rule Effectiveness (%)
596-598	PM10_RE	Real	PM10 Rule Effectiveness (%)
599-601	PM25_RE	Real	PM2.5 Rule Effectiveness (%)
602-604	NH3_RE	Real	NH ₃ Rule Effectiveness (%)

1.3.4 IDA Format Point Source Inventory

This inventory consists of a set of fixed-format detail records, preceded by a set of header records. The general format of the header records is:

```
command parm1 parm2 ... parm n
```

The command field starts in the first position. The remainder of the header is free-format, blank delimited. The specific header formats are:

```
#TYPE filetype
#COUNTRY country-name
#YEAR datayear
#DESC description
#POLID pol1 pol2 ... poln
```

The filetype and description fields are taken as the first non-blank character after the command to the last character of the record. Valid country names are US, CANADA, and MEXICO. Datayear must be a number from 1900 to 2200. The #POLID command defines the pollutants that are contained on the detail records. "pol1" is the ID of the first pollutant. "pol2" is the ID of the second pollutant, etc. Sample header records are shown below:

```
#TYPE Point Source Emission Inventory
#COUNTRY US
#YEAR 1995
#DESC Alabama 1995
#POLID CO NH3 NOX PM10 PM2_5 SO2 VOC
```

The detail records are fixed-format. The fields in each record are shown below:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-2	STID	Int	State Code
3-5	CYID	Int	County Code
6-20	PLANTID	Char	Plant Identification Code
21-35	POINTID	Char	Point Identification Code
36-47	STACKID	Char	Stack Identification Code
48-53	ORISID	Char	DOE Plant ID
54-59	BLRID	Char	Boiler Identification Code
60-61	SEGMENT	Char	DOE ID
62-101	PLANT	Char	Plant Name
102-111	SCC	Char	Source Classification Code
112-115	BEGYR	Int	Begin Year of Plant Operation
116-119	ENDYR	Int	End Year of Plant Operation
120-123	STKHGT	Real	Stack Height (ft)
124-129	STKDIAM	Real	Stack Diameter (ft)
130-133	STKTEMP	Real	Stack Gas Exit Temperature (deg F)
134-143	STKFLOW	Real	Stack Gas Flow Rate (ft ³ /sec)
144-152	STKVEL	Real	Stack Gas Exit Velocity (ft/sec)
153-160	BOILCAP	Real	Design Capacity (mmBtu/hr)
161	CAPUNITS	Char	Capacity Unit Code
162-163	WINTHRU	Real	Winter throughput (% of Annual)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
164-165	SPRTHRU	Real	Spring throughput (% of Annual)
166-167	SUMTHRU	Real	Summer throughput (% of Annual)
168-169	FALTHRU	Real	Fall throughput (% of Annual)
170-171	HOURS	Int	Normal Operating Time (hours/day)
172-173	START	Int	Normal Operation Start Time
174	DAYS	Int	Normal Operating Time (days/week)
175-176	WEEKS	Int	Normal Operating Time (weeks/year)
177-187	THRUPUT	Real	Throughput Rate (SCC units/yr)
188-199	MAXRATE	Real	Max O3 Season Rate (units/day)
200-207	HEATCON	Real	Heat Content (mmBtu/SCC unit)
208-212	SULFCON	Real	Sulfur Content (mass percent)
213-217	ASHCON	Real	Ash Content (mass percent)
218-226	NETDC	Real	Max Nameplate Capacity (MW)
227-230	SIC	Int	Standard Industrial Classification Code
231-239	LATC	Real	Latitude (decimal degrees)
240-248	LONC	Real	Longitude (decimal degrees)
249	OFFSHORE	Char	Offshore Flag
250-262	ANN1	Real	pol1 Annual Emissions (short tons)
263-275	OSD1	Real	pol1 Ozone Season Emissions (tons/day)
276-282	CE1	Real	pol1 Control Efficiency (%)
283-285	RE1	Real	pol1 Rule Effectiveness (%)
286-295	EMF1	Real	pol1 Emission Factors (SCC units)
296-298	CPRI1	Int	pol1 Primary Control Equipment Code
299-301	CSEC1	Int	pol1 Secondary Control Equipment Code
302-314	ANN2	Real	pol2 Annual Emissions (short tons)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
315-327	OSD2	Real	pol2 Ozone Season Emissions (tons/day)
328-334	CE2	Real	pol2 Control Efficiency (%)
335-337	RE2	Real	pol2 Rule Effectiveness (%)
338-347	EMF2	Real	pol2 Emission Factors (SCC units)
348-350	CPRI2	Int	pol2 Primary Control Equipment Code
351-353	CSEC2	Int	pol2 Secondary Control Equipment Code
354-366	ANN3	Real	pol3 Annual Emissions (short tons)
367-379	OSD3	Real	pol3 Ozone Season Emissions (tons/day)
380-386	CE3	Real	pol3 Control Efficiency (%)
387-389	RE3	Real	pol3 Rule Effectiveness (%)
390-399	EMF3	Real	pol3 Emission Factors (SCC units)
400-402	CPRI3	Int	pol3 Primary Control Equipment Code
403-405	CSEC3	Int	pol3 Secondary Control Equipment Code

(Repeat for n pollutants.)

1.3.5 Motor Vehicle Activity Inventory

The motor vehicle activity inventory consists of a set of free-format, space-delimited detail records preceded by a set of header records. The general format of the header records is:

```
command parm1 parm2 ... parm n
```

The command field starts in the first position. The remainder of the header is free-format, blank delimited. The specific header formats are:

```
#TYPE      filetype
#COUNTRY   country-name
#YEAR      datayear
#DESC      description
#DATA      var1 var2 ... varn
#UNITS     units1 units2 ... unitsn
```

The filetype and description fields are taken as the first non-blank character after the command to the last character of the record. Valid country names are US, CANADA, and MEXICO. Datayear must be a number from 1900 to 2200. The #DATA command defines the data contained after the source identifier on the detail records. The #UNITS command defines the units of the input data.

Sample header records are as follows:

```
#TYPE      Motor Vehicle Activity Inventory
#COUNTRY   US
#YEAR      1995
#DESC      Alabama 1995
#DATA      SPEED VMT
#UNITS     miles/hr "10E6 miles/yr" (unit, not format)
```

The detail records are free-format. The fields in each record are as follows:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	STID	Int	State Code
2	CYID	Int	County Code
3	LINKID	Char	Link Identifier (not used in SMOKE)
4	SCC	Char	Source Classification Code
5	VAR1	Real	Data for variable 1
6	VAR2	Real	Data for variable 2

(Repeat for n pollutants.)

1.3.6 Motor Vehicle Emission Inventory

The motor vehicle emission inventory consists of a set of fixed-format detail records preceded by a set of header records. The general format of the header records is:

```
command parm1 parm2 ... parm n
```

The command field starts in the first position. The remainder of the header is free-format, blank delimited. The specific header formats are:

```
#TYPE      filetype
#COUNTRY   country-name
#YEAR      datayear
#DESC      description
#POLID     pol1 pol2 ... poln
```

The filetype and description fields are taken as the first non-blank character after the command to the last character of the record. Valid country names are US, CANADA, and MEXICO. Datayear must be a number from 1900 to 2200. The #POLID command defines the pollutants that are contained on the detail records. “pol1” is the ID of the first pollutant. “pol2” is the ID of the second pollutant, etc.

Sample header records are as follows:

```
#TYPE      Motor Vehicle Emission Inventory
#COUNTRY   US
#YEAR      1995
#DESC      Alabama 1995
#POLID     CO NH3 NOX PM10 PM2_5 SO2 VOC
```

The detail records are fixed-format. The fields in each record are as follows:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-2	STID	Int	State Code
3-5	CYID	Int	County Code
6-15	LINKID	Char	Link Identifier (not used in SMOKE)
16-25	SCC	Char	Source Classification Code
26-35	ANN1	Real	pol1 Annual Emissions (short tons)
36-45	OSD1	Real	pol1 Ozone Season Emissions (tons/day)
46-55	ANN2	Real	pol2 Annual Emissions (short tons)

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
56-65	OSD2	Real	pol2 Ozone Season Emissions (tons/day)
66-75	ANN3	Real	pol3 Annual Emissions (short tons)
76-85	OSD3	Real	pol3 Ozone Season Emissions (tons/day)

(Repeat for n pollutants.)

1.4 SMOKE Input File Generation

The SMOKE Tool generates the following input files for SMOKE:

- Temporal Profiles.
- Temporal Cross-References.
- Gridding Cross-References.
- Area and Point Source Control Files.
- Speciation Cross-References.
- Pollutant File (SIPOLS).
- MOBILE Emission Factor Cross-Reference (MPLIST).

SAS is used in SMOKE Tool to generate the input files for SMOKE. The individual files contain a number of packets. Each packet contains a set of different fixed-format records. SAS routines are used for data checking, sorting, and merging the packet data into the full files.

Some additional files that are not handled in the SMOKE Tool are the ACTRACK, PCTRACK, and MPREF files. The ACTRACK and PCTRACK files identify sources to include in the control reports. These are simple, free-form ASCII files that list the counties and Source Classification Codes (SCC) for the report. Screens could have been generated to input these lists, but it is simpler for the user to type the lists directly and annotate the report node with the file name. The MPREF file defines the MOBILE5 inputs for SMOKE. It contains the statements that define SMOKE emission factor sets and imbedded MOBILE5 input statements.

The following sections describe the files processed in the SMOKE input processor of the SMOKE Tool. In most cases the SMOKE Tool allows the users to load pre-existing data, modify the data, check the packet data for data value errors, and cross-check the data in different packets.

1.4.1 Temporal Profile Files

The Temporal Profile files (ATPRO, MTPRO, and PTPRO in the SMOKE documentation) contain the monthly, weekly, and diurnal temporal profiles used to convert the annual inventory emissions to hourly pollutant emissions. The files are all in the same format. Each file is an ASCII fixed-format file that can contain four types of profiles. Each profile type is contained in a

packet. The packet begins with a packet identifier (/identifier/) and ends with a packet end statement (/END/). The temporal factors entered are relative weights and do not have to sum to one. The SMOKE Tool will internally convert the weights that sum to one. The packets in file order are:

<u>Identifier</u>	<u>Description</u>
/MONTHLY/	Monthly temporal weighting factors
/WEEKLY/	Weekly temporal weighting factors
/DIURNAL WEEKDAY/	Weekday diurnal temporal weighting factors
/DIURNAL WEEKEND/	Weekend diurnal temporal weighting factors

The weekend diurnal factors are optional. At least one monthly profile is required. The monthly temporal weighting factor record format is:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-5	MPROF	Int	Monthly temporal profile code
7-9	JAN	Int	Temporal weight for January
11-13	FEB	Int	Temporal weight for February
15-17	MAR	Int	Temporal weight for March
.			
.			
.			
51-53	DEC	Int	Temporal weight for December
54-58	MSUM	Int	Total of all weights for year

The weekly temporal weighting factors record format is:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-5	WPROF	Int	Weekly temporal profile code
7-9	MON	Int	Temporal weight for Monday

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
11-13	TUE	Int	Temporal weight for Tuesday
15-17	WED	Int	Temporal weight for Wednesday
.			
.			
.			
31-33	SUN	Int	Temporal weight for Sunday
34-38	WSUM	Int	Total of all weights for week

The diurnal temporal weighting factor record format is:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-5	DPROF	Int	Diurnal temporal profile code
7-9	WGT1	Int	Temporal weight for 00:00-01:00
11-13	WGT2	Int	Temporal weight for 01:00-02:00
15-17	WGT3	Int	Temporal weight for 02:00-03:00
.			
.			
.			
99-101	WGT24	Int	Temporal weight for 23:00-24:00
103-105	DSUM	Int	Total of all weights for day

1.4.2 Temporal Cross-Reference Files

The Temporal Cross-Reference files (listed as ATREF, MTREF, and PTREF in the SMOKE documentation) match emission sources to the profiles defined in the Temporal Profile files. The files are all ASCII fixed-format files.

The area source temporal cross-reference file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	SCC	Int	Source Classification Code (10 or 4 digits)
2	MPROF	Int	Monthly temporal profile code
3	WPROF	Int	Weekly temporal profile code
4	DPROF	Int	Diurnal temporal profile code (used for both weekday and weekend)
5	COUNTY	Int	Country/State/County Code

The combinations of keys allowed for the area source temporal cross-reference (in search order) are:

- county, scc (10 digits).
- county, scc (4 digits).
- state, scc (10 digits).
- state, scc (4 digits).
- country, scc (10 digits).
- country, scc (4 digits).
- scc (10 digits).
- scc (4 digits).
- county.
- state.
- country.

Countries, states, and counties are represented in the COUNTY field in the following manner:

- n00000 Country. All states and counties in country n. An important note is that the U.S. country number is zero.
- nss000 State. All counties in country n, state ss.
- nssccc County. Country n, state ss, county ccc.

The point source temporal cross-reference file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	SCC	Char	Source Classification Code (8 or 3 digits)
2	MPROF	Int	Monthly temporal profile code
3	WPROF	Int	Weekly temporal profile code
4	DPROF	Int	Diurnal temporal profile code (used for both weekday and weekend)
5	COUNTY	Int	Country/State/County Code
6	PLANTID	Char	Plant ID
7	POINTID	Char	Point ID
8	STKID	Char	Stack ID
9	SEGMENT	Char	Segment Number

The combinations of keys allowed for the point source temporal cross-reference (in search order) consist of two levels: plant and non-plant. The plant-level key combinations can be any of the following with an eight-digit SCC, a three-digit SCC, or no SCC:

- county, plantid, pointid, stkid, segment, scc.
- county, plantid, pointid, stkid, segment.
- county, plantid, pointid, stkid.
- county, plantid, pointid.
- county, plantid.

The non-plant level key combinations can be:

- county, scc (eight digits).
- county, scc (three digits).
- state, scc (eight digits).
- state, scc (three digits).
- country, scc (eight digits).
- country, scc (three digits).
- scc (eight digits).
- scc (three digits).
- county.
- state.

- country.

The mobile source temporal cross-reference file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	SCC	Char	Source Classification Code (10 digits) digits 3-4 are roadclass digits 5-7 are vehtype digits 8-10 are process
2	MPROF	Int	Monthly temporal profile code
3	WPROF	Int	Weekly temporal profile code
4	DPROF	Int	Diurnal temporal profile code (used for both weekday and weekend)
5	COUNTY	Int	Country/State/County Code
6	LINKID	Int	Link ID code (blank or -9 for non-link record)

The combinations of keys allowed for the mobile source temporal cross-reference are:

- county, linkid, roadclas, vehtype, process.
- county, linkid, roadclas, vehtype.
- county, linkid, roadclas.
- county, roadclas, vehtype, process.
- county, roadclas, vehtype.
- county, roadclas.
- state, roadclas, vehtype, process.
- state, roadclas, vehtype.
- state, roadclas.
- state, vehtype, process.
- state, vehtype.
- state.
- roadclas, vehtype, process.
- roadclas, vehtype.
- roadclas.
- vehtype, process.
- vehtype.
- default.

The SCC subfields mentioned in the above list are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
3-4	VEHTYPE	Int	Vehicle Type
5-7	ROADCLAS	Int	Road Class
8-10	PROCESS	Char	Emissions Process

1.4.3 Gridding Cross-Reference Files

The Gridding Cross-Reference files (AGREF, BGREF, and MGREF in the SMOKE documentation) match emission sources to gridding spatial surrogates. If necessary, the SMOKE Tool will generate the Emission Modeling System (EMS)-95-derived cross-reference formats. These are free-format, space-delimited ASCII files. The file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code
3	SSC	Int	Spatial Surrogate Code

1.4.4 Area and Point Source Control Files

The Control files (GCNTL in the SMOKE documentation) define control and projection information for a SMOKE case. The files are basically in the same format. Both are ASCII files that can contain packets defining the various types of control and projection information. The packets begin with a packet identifier (/identifier/) and end with a packet end statement (/END/). Between the identifier and end statement is a set of records containing the packet data. The packets in file order are:

<u>Identifier</u>	<u>Description</u>
/CTG/	Control technology guideline controls
/CONTROL/	Control factor specifications
/ALLOWABLE/	Allowable emission limits
/REACTIVITY/	Reactivity

<u>Identifier</u>	<u>Description</u>
/PROJECTION by fy/	Projection factors for area or point sources. "by" and "fy" are four-digit years; "by" is the from year and "fy" is the to year for the projection factors.

The data records in each packet consist of a single record type. All records are free-format. Generally, area source data will have a 10-digit SCC. Point Source files use either a 4-digit SIC or an 8-digit SCC.

The Control Technology Guideline (CTG) controls record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code (10-digit for area, 8-digit for point)
3	POLID	Char	Pollutant ID
4	CTG	Real	CTG Control Factor
5	CTGCUT	Real	Emission Cutoff for Application of Control (tons/day)
6	MACT	Real	Maximum Achievable Control Technology Control Factor
7	RACT	Real	Reasonably Achievable Control Technology Control Factor

The control packet record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code (10-digit for area, 8-digit for point)
3	POLID	Char	Pollutant ID
4	CPRI	Int	Primary Control Equipment Code (blank or zero applies to all equipment)

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
5	CEEF	Real	Control Efficiency (0-100%)
6	RLEF	Real	Rule Effectiveness (0-100%)
7	RLPN	Real	Rule Penetration Rate (0-100%)
8	SIC	Int	Standard Industrial Category (only for point)
9	PLANTID	Char	Plant ID (only for point)
10	POINTID	Char	Point ID (only for point)
11	STKID	Char	Stack ID (only for point)
12	SEGMENT	Char	Segment (only for point)

The allowable emissions record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code (10-digit for area, 8-digit for point)
3	POLID	Char	Pollutant ID
4	CTLFAC	Real	Control Factor
5	CAP	Real	Allowable Emissions (Cap)
6	REP	Real	Allowable Emissions (Replace)
7	SIC	Int	Standard Industrial Category (only for point)
8	PLANTID	Char	Plant ID (only for point)
9	POINTID	Char	Point ID (only for point)
10	STKID	Char	Stack ID (only for point)
11	SEGMENT	Char	Segment (only for point)

The reactivity packet record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code (10-digit for area, 8-digit for point)
3	POLID	Char	Pollutant ID
4	REP	Real	Allowable Emissions (Replace)
5	RF	Real	Projection Factor
6	NEWSCC	Int	New Source Classification Code
7	INPRF	Int	Speciation Profile Number
8	MKTPEN	Char	Market Penetration Rate of New Speciation
9	SIC	Int	Standard Industrial Category (only for point)
10	PLANTID	Char	Plant ID (only for point)
11	POINTID	Char	Point ID (only for point)
12	STKID	Char	Stack ID (only for point)
13	SEGMENT	Char	Segment (only for point)

The projection record format is:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	SCC	Int	Source Classification Code (10-digit for area, 8-digit for point)
3	RF	Real	Projection Factor
4	SIC	Int	Standard Industrial Category (only for point)

1.4.5 Speciation Cross-Reference File

The Speciation Cross-Reference file is an ASCII file that maps pollutant sources to the Speciation profiles defined for pollutants. The fields on each record are:

<u>Position</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1-10	SCC	Char	Source Classification Code
12-16	INPRF	Char	Profile Number
18-33	POLID	Char	Pollutant ID
35-40	COUNTY	Int	Country/State/County Code
42-56	PLANTID	Char	Plant ID (only for point source)
58-72	POINTID	Char	Point ID (only for point source)

The valid source definition field combinations can be:

- country, std, cyid, plantid, pointid, scc.
- country, std, cyid, plantid, scc.
- country, std, cyid, scc.
- country, std, scc.
- country, scc.
- scc.
- country, std, cyid (county level default).
- country, std (state level default).
- country (country level default).
- <none> (domain level default).

Countries, states, and counties are represented in the COUNTY field in the following manner:

- n00000 Country. All states and counties in country n. Please note that the country number for the U.S. is zero.
- nss000 State. All counties in country n, state ss.
- nssccc County. Country n, state ss, county ccc.

1.4.6 Pollutant Definition Files

The SMOKE Input Pollutant Definition files (SIPOLS in the SMOKE documentation) match SMOKE internal pollutant numbers to pollutant IDs. It also selects the pollutants that SMOKE will process. Pollutants will appear in the order entered in SMOKE reports. These are

free-format, space-delimited ASCII files. The file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	POLNUM	Int	Internal Pollutant Number
2	POLID	Char	Pollutant ID

1.4.7 MOBILE Emission Factor Cross-Reference Files

The MOBILE Emission Factor Cross-Reference files (MPLIST in the SMOKE documentation) assigns mobile emission factors to roads in individual counties, states, or countries. These are free-format, space-delimited ASCII files. The file record format is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Country/State/County Code
2	ROADCLAS	Int	Road Class code (zero for default)
3	LINKID	Int	Link ID (zero if not link-specific)
4	PSI1	Int	PSI Entry 1
5	PSI2	Int	PSI Entry 2
6	PSI3	Int	PSI Entry 3
.			
.			
.			
27	PSI24	Int	PSI Entry 24

A PSI is the identifier specified for an emission factor set in the SMOKE MOBILE5 input file (MPREF). PSI references must be specified for each of the 24 hours of the day. In the typical case, the emission factors (and therefore the PSIs) do not vary by day. The most basic way to structure this file is to put one PSI in each entry. A shortcut format can also be used to indicate blocks of hours that use a specific PSI. This format is the number of hours, followed by an asterisk, followed by a PSI. Sample shortcut entries are shown below:

<u>PSI Entries</u>	<u>Description</u>
24*001	PSI 001 for all 24 hours
8*003 8*004 8*001	PSI 003 for hours 1 to 8 PSI 004 for hours 9 to 16 PSI 001 for hours 17 to 24
001 6*002 8*003 004 8*005	PSI 001 for hour 1 PSI 002 for hours 2 to 7 PSI 003 for hours 8 to 15 PSI 004 for hour 16 PSI 005 for hours 17 to 24

1.4.8 EGAS Files

The growth factor files included with Models-3 are a compact form of the EGAS growth factor files. Each file consists of the following sections in the order shown:

- Comments. These may be in any format.
- Growth factor profiles. This section is preceded by a header record that specifies the base year for the profiles and the last year of growth factors in the profiles. The first year in the profile is the base year plus one. The header record is followed by as many profile records as necessary. The format of the header record is:

#EGASPROF baseyr lastyr

- Growth factor cross-references. This section associates sources with growth profiles by geographic areas and source classification codes. It is preceded by a header record. The header record is followed by as many profile records as necessary. The format of the header record is:

#EGASXREF

The format of the profile records is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	PROF	Int	Profile Number
2	GF1	Real	Growth factor, 1 st year (baseyr+1)

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
3	GF1	Real	Growth factor, 2 nd year (baseyr+2)
4	GF1	Real	Growth factor, 3 rd year (baseyr+3)
.			
.			
.			
n+1	GF1	Real	Growth factor, n th year (lastyr)

The format of the cross-reference records is:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	COUNTY	Int	Profile Number
2	XSCC	Char	Source Classification Code. Prefixed by source type (A=Area, P=Point, M=Mobile)
3	PROF	Int	Profile Number

1.5 Process for Registering Datasets from within the SMOKE Tool

For SMOKE to use the files created by the SMOKE Tool, it is easiest to register these datasets via the Dataset Manager. Registering the datasets is an option for the user when running the SMOKE Tool. After registering a dataset, the SMOKE Tool will still be able to overwrite the contents of the file registered, so a new dataset does not need to be registered every time.

2.0 STUDY PLANNER

Changes were made to the Study Planner to get the necessary information to SMOKE. The information that SMOKE must get at the time of the execution of the program is:

- Grid Information (including horizontal grid, coordinate, and vertical layer information).
- Case Information.
- Chemical Mechanism Information.
- Extracted Continuous Emission Monitoring (CEM) Data for Case.
- File Lists for SMOKE Input.

2.1 Grid Information

In the past, Study Planner did not allow selection of grid information because Model Builder compiled the grid information directly into the FORTRAN code. This caused the program to require compiling every time the grid was changed. To implement SMOKE and other programs so that the executables do not require recompiling every time, the grid information from the Science Manager was made available to the SMOKE program. The horizontal grid, the coordinate, and the vertical layer information are written into one file. An environment variable whose value is the location of the file is also created. For each variable, the file contains a Record 1 for the variable description and a Record 2 for the variable values. However, there is only a Record 1 (no Record 2) when the variable has a single value. In this case, DIM is set to zero, and the variable value follows DIM on Record 1, space-delimited. The format of the file looks like the following:

Record 1:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	KEYWORD	Char	Keyword for variable name
2	DIM	Int	Number of dimensions
3	SIZE1	Int	Number of elements for first dimension
4	SIZE2	Int	Number of elements for second dimension
5	SIZE3	Int	Number of elements for third dimension

(Repeat for each dimension.)

Record 2:

All values of the variable, space-delimited. The first dimension is incremented first, then the second, and so forth. The following keywords are used:

- GDNAME_GD Grid Name
- GDDDESC_GD Grid Description
- GDTYP_GD Coordinate System Type:
 - LATGRD3 Lat/Lon Coordinates (unused)
 - LAMGRD3 Lambert Coordinates
 - MERGRD3 Mercator Coordinates
 - STEGRD3 Stereographic Coordinates
 - UTMGRD3 UTM Coordinates
- GDUNT_GD Grid Units (DEGREES)
- P_ALP_GD Alpha Value (different for each coordinate system type, see below)

- P_BET_GD Beta Value
- P_GAM_GD Gamma Value
- P_XCENT_GD X Center, $-180 < P_XCENT_GD \leq 180$, for the center of the grid's
respective Cartesian coordinate system
- P_YCENT_GD Y Center, $-90 \leq Y \leq 90$, for the center of the grid's respective
Cartesian coordinate system
- XORIG_GD X Origin. For Lambert, Mercator, UTM, and Stereographic,
XORIG_GD and YORIG_GD are locations in map units (km) of
the origin cell (1,1) (lower left corner) of the horizontal grid
measured from (P_XCENT_GD, P_YCENT_GD). For Lat/Lon,
units are degrees.
- YORIG_GD Y Origin
- XCELL_GD Delta X Cell Size
- YCELL_GD Delta Y Cell Size
- NCOLS Column Size
- NROWS Row Size
- NTHIK External Boundary Thickness
- NLAYS Number of Vertical Layers
- VG_TYP_GD Vertical Coordinate Type:
 - VSGSPH3 sigma-P Hydrostatic
 - VSGSPN3 sigma-P Non Hydrostatic
 - VGSIGZ3 sigma-Z
 - VGPRES3 Pressure Coordinates
 - VGZVAL3 for Z (meters above mean sea level)
 - VGHVAL3 for H (meters above ground)
 - IMISS3 for vertical coordinates not stored in VGLVSD_GD (e.g.,
temporally or spatially changing vertical coordinates)
- VGTPUN_GD Units of Vertical Coordinate Top
- VGTOP_GD Value of Model Top
- VGLVS_GD List of Vertical Coordinate Surface Values

The definitions of the map projection specification parameters:

- P_ALP_GD (PROJ_ALPHA)
- P_BET_GD (PROJ_BETA)
- P_GAM_GD (PROJ_GAMMA)

depend upon the projection type, as follows:

- Lambert $P_ALP_GD \leq P_BET_GD$ are the two latitudes that determine the projection
cone; P_GAM_GD is the central meridian.

- Mercator P_ALP_GD and P_BET_GD are the latitude and longitude of the coordinate origin (within the tangent circle); P_GAM_GD is the angle between the cylinder axis and the North polar axis.
- Stereographic P_ALP_GD and P_BET_GD are the latitude and longitude of the point of tangency; P_GAM_GD is the angle from true North to the Y-axis.
- UTM P_ALP_GD is the UTM zone. P_BET_GD and P_GAM_GD are unused.
- lat/lon is currently not used. Coordinate units are degrees, with $-180.0 < X \leq 180.0$, $-90.0 \leq Y < 90.0$.
- If P_ALP_GD < -9E36, then the grid description is missing or invalid.

The environment variable with the full path location of the grid file is G_GRIDPATH, and this file contains all the necessary grid information. The grid and vertical layer is selected from pick lists at the plan and node level, as is done for case definition. Also, a selection for all three of these (horizontal grid, vertical layer, and case) was added to the study level so there can be a default for the entire study. The file is created every time before the node is executed. It is created in the same directory where the Study Planner saves the script and log files. This is also the default location for all the output files. An environment variable, STUDY_PATH, which has this path as its value, was added.

The following is an example of G_GRIDPATH:

```
GDTYP_GD 0 LAMGRD3
P_ALP_GD 0 30.000000
P_BET_GD 0 60.000000
P_GAM_GD 0 -90.000000
P_XCENT_GD 0 -90.000000
P_YCENT_GD 0 40.000000
VGTYPE_GD 0 VGSGPN3
VGTOP_GD 0 10000.0
VGTPUN_GD 0 Pa
VGLVUN_GD 0 none
VGLVS_GD 1 16
1.0 0.995 0.985 0.970 0.950 0.920 0.870 0.800 0.700 0.600 0.500 0.400 0.300 0.200
0.100 0.00
X3FACE_GD 1 16
0 0.005 0.015 0.03 0.05 0.08 0.13 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
GDNAME_GD 0 c_ctm_36_demo
XORIG_GD 0 -8.640000E+05
YORIG_GD 0 -1.512000E+06
XCELL_GD 0 3.600000E+04
YCELL_GD 0 3.600000E+04
GL_NROWS 0 69
GL_NCOLS 0 75
NPCOL 0 1
NPROW 0 1
```

```

NROWS 0 69
NCOLS 0 75
NTHIK 0 1
MNDIS 0 2
MEDIS 0 2
MSDIS 0 2
MWDIS 0 2
NLAYS 0 15

```

2.2 Case Information

Currently, Study Planner sends the case information in environment variables. There are four environment variables as follows:

- G_STDATE Julian Date in YYYYDDD format.
- G_STTIME Time in HHMMSS format.
- G_TSTEP Time step in HHMMSS format.
- G_RUNLEN Run length in HHMMSS format.

YYYY is the four-digit year. DDD is the number of days since the beginning of the year. HH is the hours (this does not have to be two digits; it can be any number of digits). MM is the minutes (two digits). SS is the seconds (two digits).

This is the only case information necessary for the SMOKE program. The following is an example of the Case environment variables:

```

G_RUNLEN=240000
G_STDATE=1995192
G_STTIME=000000
G_TSTEP=10000

```

2.3 Chemical Mechanism Information

Chemical mechanism information is passed to SMOKE as it is executed in Study Planner. This is done using the same idea discussed in Section 2.1, Grid Information. The files are created, and environment variables provide the location of these files. In addition, CHEMMECH outputs the RXCM.EXT and RXDT.EXT files into one file in the same format as was discussed in Section 2.1. The environment variable name is C_RX for the file created by CHEMMECH and C_TABLE for the remaining chemical mechanism files. The chemical mechanism is a selection at the study and plan levels.

The C_TABLE environment variable provides the path to a file that combines all the variables provided by the files beginning with AE_, GC_, NR_, TR_. These files are created from the tables found in the Chemical Mechanism of the Science Manager. The GC_ files include the Gas

data, the AE_ files include the Aerosol data, the NR_ files include the Non-Reactive data, and the TR_ files include the Tracer data. The C_TABLE file contains the following keywords, where xx is the two-letter code (AE, GC, NR, TR) and ? is the one-letter code (A, G, N, T):

N_xx_?2AE	Number of this type
xx_?2AE	List of species
xx_?2AE_MAP	Surrogate mapping for Aerosol conversion
xx_?2AE_FAC	Factor
N_xx_?2AQ	Number of this type
xx_?2AQ	List of species
xx_?2AQ_MAP	Surrogate mapping for Aqueous conversion
xx_?2AQ_FAC	Factor
N_xx_ADV	Number of this type
xx_ADV	List of species
xx_ADV_MAP	List of species index (row in table) to have species advected
N_xx_CONC	Number of this type
xx_CONC	List of species
xx_CONC_MAP	List of species index (row in table) to have concentration outputted
N_xx_DDEP	Number of this type
xx_DDEP	List of species
xx_DDEP_MAP	List of species index (row in table) to have dry deposition outputted
N_xx_DEPV	Number of this type
xx_DEPV	List of species
xx_DEPV_MAP	Surrogate mapping for deposition velocity
xx_DEPV_FAC	Factor
N_xx_DIFF	Number of this type
xx_DIFF	List of species
xx_DIFF_MAP	List of species index (row in table) to have species diffused
N_xx_EMIS	Number of this type
xx_EMIS	List of species
xx_EMIS_MAP	Surrogate mapping for emissions processing
xx_EMIS_FAC	Factor
N_xx_ICBC	Number of this type
xx_ICBC	List of species
xx_ICBC_MAP	Surrogate mapping for initial and boundary condition processing
xx_ICBC_FAC	Factor
N_xx_SCAV	Number of this type
xx_SCAV	List of species
xx_SCAV_MAP	Surrogate mapping for Henry's Law constant
xx_SCAV_FAC	Factor
N_xx_SPC	Number of this type
xx_SPC	List of species

xx_MOLWT	List of molecular weights
N_xx_WDEP	Number of this type
xx_WDEP	List of species
xx_WDEP_MAP	List of species index (row in table) to have wet deposition outputted

In addition, any column added to this table that ends in _SUR or _FAC also adds variables for the matching type (_SUR creates the MAP and _FAC creates the FAC). Since the N_ (number of) variable previously was in each individual file, it could have a different value for the surrogate and the factor list. Now both lists need to be the same length, and the system enters the defaults (species for MAP and 1.0 for FAC) for any missing values.

The following is an example of the C_TABLE file:

```

N_GC_SPC 0 36
N_GC_SPCD 0 37
GC_SPC 1 36
NO2 NO O O3 NO3 O1D OH HO2 N2O5 HNO3 HONO PNA H2O2 CO FORM ALD2 C2O3 XO2 PAN PAR
XO2N ROR NTR OLE ETH TOL CRES TO2 OPEN CRO XYL MGLY ISOP ISPD SO2 SULF
GC_MOLWT 1 36
46.0 30.0 16.0 48.0 62.0 16.0 17.0 33.0 108.0 63.0 47.0 79.0 34.0 28.0 30.0 44.0
75.0 1.0 121.0 14.0 1.0 31.0 130.0 27.0 28.0 92.0 108.0 109.0 84.0 139.0 106.0
72.0 68.0 70.0 64.0 98.0
N_GC_EMIS 0 13
N_GC_EMISD 0 13
GC_EMIS 1 13
NO2 NO CO FORM ALD2 PAR OLE ETH TOL XYL ISOP SO2 SULF
GC_EMIS_MAP 1 13
1 2 14 15 16 20 24 25 26 31 33 35 36
GC_EMIS_FAC 1 13
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
N_GC_ICBC 0 0
N_GC_ICBCD 0 1
GC_ICBC 1 0
GC_ICBC_MAP 1 0
GC_ICBC_FAC 1 0
N_GC_DEPV 0 9
N_GC_DEPVD 0 9
GC_DEPV 1 9
VD_NO2 VD_NO VD_O3 VD_HNO3 VD_H2O2 VD_HCHO VD_ALD VD_SO2 VD_SULF
GC_DEPV_MAP 1 9
1 2 4 10 13 15 16 35 36
GC_DEPV_FAC 1 9
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
N_GC_G2AE 0 0
N_GC_G2AED 0 1
GC_G2AE 1 0
GC_G2AE_MAP 1 0
GC_G2AE_FAC 1 0
N_GC_G2AQ 0 0
N_GC_G2AQD 0 1
GC_G2AQ 1 0
GC_G2AQ_MAP 1 0
GC_G2AQ_FAC 1 0
N_GC_SCAV 0 0

```

```

N_GC_SCAVD 0 1
GC_SCAV 1 0
GC_SCAV_MAP 1 0
GC_SCAV_FAC 1 0
N_GC_ADV 0 36
N_GC_ADV_D 0 36
GC_ADV 1 36
NO2 NO O O3 NO3 O1D OH HO2 N2O5 HNO3 HONO PNA H2O2 CO FORM ALD2 C2O3 XO2 PAN PAR
XO2N ROR NTR OLE ETH TOL CRES TO2 OPEN CRO XYL MGLY ISOP ISPD SO2 SULF
GC_ADV_MAP 1 36
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36
N_GC_DIFF 0 36
N_GC_DIFF_D 0 36
GC_DIFF 1 36
NO2 NO O O3 NO3 O1D OH HO2 N2O5 HNO3 HONO PNA H2O2 CO FORM ALD2 C2O3 XO2 PAN PAR
XO2N ROR NTR OLE ETH TOL CRES TO2 OPEN CRO XYL MGLY ISOP ISPD SO2 SULF
GC_DIFF_MAP 1 36
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36
N_GC_CONC 0 36
N_GC_CONC_D 0 36
GC_CONC 1 36
NO2 NO O O3 NO3 O1D OH HO2 N2O5 HNO3 HONO PNA H2O2 CO FORM ALD2 C2O3 XO2 PAN PAR
XO2N ROR NTR OLE ETH TOL CRES TO2 OPEN CRO XYL MGLY ISOP ISPD SO2 SULF
GC_CONC_MAP 1 36
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36
N_GC_DDEP 0 9
N_GC_DDEP_D 0 9
GC_DDEP 1 9
NO2 NO O3 HNO3 H2O2 FORM ALD2 SO2 SULF
GC_DDEP_MAP 1 9
1 2 4 10 13 15 16 35 36
N_GC_WDEP 0 0
N_GC_WDEP_D 0 1
GC_WDEP 1 0
GC_WDEP_MAP 1 0
N_AE_SPC 0 0
N_AE_SPC_D 0 1
AE_SPC 1 0
AE_MOLWT 1 0
N_AE_EMIS 0 0
N_AE_EMIS_D 0 1
AE_EMIS 1 0
AE_EMIS_MAP 1 0
AE_EMIS_FAC 1 0
N_AE_ICBC 0 0
N_AE_ICBC_D 0 1
AE_ICBC 1 0
AE_ICBC_MAP 1 0
AE_ICBC_FAC 1 0
N_AE_DEPV 0 0
N_AE_DEPV_D 0 1
AE_DEPV 1 0
AE_DEPV_MAP 1 0
AE_DEPV_FAC 1 0
N_AE_A2AQ 0 0
N_AE_A2AQ_D 0 1

```

```

AE_A2AQ 1 0
AE_A2AQ_MAP 1 0
AE_A2AQ_FAC 1 0
N_AE_SCAV 0 0
N_AE_SCAVD 0 1
AE_SCAV 1 0
AE_SCAV_MAP 1 0
AE_SCAV_FAC 1 0
N_AE_ADV 0 0
N_AE_ADVD 0 1
AE_ADV 1 0
AE_ADV_MAP 1 0
N_AE_DIFF 0 0
N_AE_DIFFD 0 1
AE_DIFF 1 0
AE_DIFF_MAP 1 0
N_AE_CONC 0 0
N_AE_CONCD 0 1
AE_CONC 1 0
AE_CONC_MAP 1 0
N_AE_DDEP 0 0
N_AE_DDEPD 0 1
AE_DDEP 1 0
AE_DDEP_MAP 1 0
N_AE_WDEP 0 0
N_AE_WDEPD 0 1
AE_WDEP 1 0
AE_WDEP_MAP 1 0
N_NR_SPC 0 0
N_NR_SPCD 0 1
NR_SPC 1 0
NR_MOLWT 1 0
N_NR_EMIS 0 0
N_NR_EMISD 0 1
NR_EMIS 1 0
NR_EMIS_MAP 1 0
NR_EMIS_FAC 1 0
N_NR_ICBC 0 0
N_NR_ICBCD 0 1
NR_ICBC 1 0
NR_ICBC_MAP 1 0
NR_ICBC_FAC 1 0
N_NR_DEPV 0 0
N_NR_DEPVD 0 1
NR_DEPV 1 0
NR_DEPV_MAP 1 0
NR_DEPV_FAC 1 0
N_NR_N2AE 0 0
N_NR_N2AED 0 1
NR_N2AE 1 0
NR_N2AE_MAP 1 0
NR_N2AE_FAC 1 0
N_NR_N2AQ 0 0
N_NR_N2AQD 0 1
NR_N2AQ 1 0
NR_N2AQ_MAP 1 0
NR_N2AQ_FAC 1 0
N_NR_SCAV 0 0
N_NR_SCAVD 0 1

```

```

NR_SCAV 1 0
NR_SCAV_MAP 1 0
NR_SCAV_FAC 1 0
N_NR_ADV 0 0
N_NR_ADVD 0 1
NR_ADV 1 0
NR_ADV_MAP 1 0
N_NR_DIFF 0 0
N_NR_DIFFD 0 1
NR_DIFF 1 0
NR_DIFF_MAP 1 0
N_NR_CONC 0 0
N_NR_CONCD 0 1
NR_CONC 1 0
NR_CONC_MAP 1 0
N_NR_DDEP 0 0
N_NR_DDEPD 0 1
NR_DDEP 1 0
NR_DDEP_MAP 1 0
N_NR_WDEP 0 0
N_NR_WDEPD 0 1
NR_WDEP 1 0
NR_WDEP_MAP 1 0
N_TR_SPC 0 0
N_TR_SPCD 0 1
TR_SPC 1 0
TR_MOLWT 1 0
N_TR_EMIS 0 0
N_TR_EMISD 0 1
TR_EMIS 1 0
TR_EMIS_MAP 1 0
TR_EMIS_FAC 1 0
N_TR_ICBC 0 0
N_TR_ICBCD 0 1
TR_ICBC 1 0
TR_ICBC_MAP 1 0
TR_ICBC_FAC 1 0
N_TR_DEPV 0 0
N_TR_DEPVD 0 1
TR_DEPV 1 0
TR_DEPV_MAP 1 0
TR_DEPV_FAC 1 0
N_TR_T2AE 0 0
N_TR_T2AED 0 1
TR_T2AE 1 0
TR_T2AE_MAP 1 0
TR_T2AE_FAC 1 0
N_TR_T2AQ 0 0
N_TR_T2AQD 0 1
TR_T2AQ 1 0
TR_T2AQ_MAP 1 0
TR_T2AQ_FAC 1 0
N_TR_SCAV 0 0
N_TR_SCAVD 0 1
TR_SCAV 1 0
TR_SCAV_MAP 1 0
TR_SCAV_FAC 1 0
N_TR_ADV 0 0
N_TR_ADVD 0 1

```

```

TR_ADV 1 0
TR_ADV_MAP 1 0
N_TR_DIFF 0 0
N_TR_DIFFD 0 1
TR_DIFF 1 0
TR_DIFF_MAP 1 0
N_TR_CONC 0 0
N_TR_CONCD 0 1
TR_CONC 1 0
TR_CONC_MAP 1 0
N_TR_DDEP 0 0
N_TR_DDEPD 0 1
TR_DDEP 1 0
TR_DDEP_MAP 1 0
N_TR_WDEP 0 0
N_TR_WDEPD 0 1
TR_WDEP 1 0
TR_WDEP_MAP 1 0

```

2.4 Extracting CEM Data for Case

A specific SAS program was developed and is executed as a node in a plan. The program has two inputs:

- The Models-3 case description annotated to the node via environment variables G_STDATE, G_STTIME, and G_RUNLEN.
- A CEM dataset (SAS) specified through an input link. The user has to annotate the link, selecting a registered CEM dataset that contains data for the case period.

The program extracts all CEM data for the case time period, sorting the data into the sequence needed by SMOKE. The output is a free-form, space-delimited ASCII file. The data fields in each record are:

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
1	ORIS	Int	DOE Plant ID
2	BLRID	Char	Boiler Identification Code
3	DATE	Int	Measurement Date (yyyymmdd)
4	TIME	Int	Measurement Time (hhmmss)
5	HTINPUT	Real	Heat Input (10**6 BTU) (units, not format)
6	NOXRATE	Real	NOX Emission Rate (stoichiometric data)
7	SO2RATE	Real	SO ₂ Emission Rate (lb/hr)

<u>Field</u>	<u>Name</u>	<u>Type</u>	<u>Description</u>
8	CO2RATE	Int	CO2 Emission Rate (lb/hr)

The NOX emission rate (NOX_RATE) in lb/hr is computed using the following equation:

$$\text{NOX_RATE} = \text{NOXRATE} * \text{HTINPUT}$$

The program writes the output file to a file described in an output link specification.

2.5 File Lists for SMOKE Input

The current SMOKE implementation expects files that contain lists of other files to process, a specific case being the inventory files. Inventory files are merged into one file, as described previously in this document.

APPENDIX A

Running SMOKE Tool Outside of Models-3

Smoke Tool may be operated independent of the Models-3 framework. However, SAS 8 and Arc/Info 8 must be present and available to SMOKE Tool (using a common directory or pointers) and the Models-3 directory structure (environment) must be established. Except for establishing the user-defined grid and gridding spatial surrogates, SMOKE Tool may be operated through its SAS windows in the same manner as through the Models-3 framework by executing the executable of the desired portion of the SMOKE Tool from the binary directory where the programs are installed. In order to define and create grid and gridded surrogate files without using the Models-3 Study Planner windows, alternative procedures using scripts must be used.

SMOKETOOL uses the following scripts:

\$M3MSTOOL/bin/smoktool_grid	Spatial Surrogate Processor
\$M3MSTOOL/bin/smoktool_ida	Inventory Data Analyzer
\$M3MSTOOL/bin/smoktool_in	SMOKE Input File Generator
\$M3MSTOOL/bin/smoktool_spec	Emission Chemical Mechanism

All of the scripts need the Models-3 environment setup. In other words, the directory and file structure used by Models-3 is also necessary to run the SMOKE Tool using scripts. The primary items needed from the Models-3 setup are M3MSTOOL, sasroot, ARCHOME, and ARCBIN. The Inventory Data Analyzer and SMOKE Input File Generator need no other setup. The Emission Chemical Mechanism must have the following environment variables set before execution:

EMISMECH	Name of mechanism
EMISMECH_TYPE	Type of mechanism (CB4, RADM2, NONE)
M3DATA	Models-3 data base directory

The spatial surrogate processor is more complicated. The following sections are a guide to the procedure:

smoktool_grid_instr	Instructions on running the script
test_surrogates	Test script that runs all processor steps
m3grid_c_cmaqmet_36eval	Sample grid description file

The test script was setup for a local test environment, and some directories must be changed to the users environment (those with susick in the name).

Smoketool grid instructions

The spatial surrogate processor has three steps:

define	Defines the grid
coverages	Generates gridded coverages and data files in the grid directories
surrogates	Computes spatial surrogates from the gridded data computed in the coverages step

The spatial surrogate processor was designed run as a node in a study plan. The Models-3 environment must be setup prior to executing any of the step in a stand-alone mode. Once the environment has been established, the following UNIX commands will execute the steps:

```
$M3MSTOOL/bin/smoktool_grid  define
$M3MSTOOL/bin/smoktool_grid  coverages
$M3MSTOOL/bin/smoktool_grid  surrogates
```

The base environment variables used by all the steps are:

EMS_HOME	Workspace path. Base directory for gridding.
HPATHNAME	Models-3 grid name.
GISDB	GIS data base directory
STUDY_PATH	Directory where program log and report will be written

EMS_HOME and HPATHNAME are used to determine the grid base directory:

```
$EMS_HOME/gridspec/$HPATHNAME
```

The standard GISDB setting is \$M3DATA/nostudies/gisdb. STUDY_PATH is normally supplied by Models-3, but must be specified in the stand-alone mode. Two files will be written to \$STUDY_PATH; program.log and program.lst.

ARC logs may be saved in a file by redirecting the script output to a file.

Define Grid

Define defines the grid. It will:

Create the grid directories
 Input the grid description
 Create the initial coverages for the grid

The only input file to the step is defined by the following environment variable:

G_GRIDPATH Name of Models-3 grid description file. Supplied by Models-3 in study planner. File must be in Models-3 format.

The sub-directories created under the grid base directory will be:

sas Gridded data directory. Contains grid description and gridded data files from coverages

gis Gridded GIS directory. Contains copy of ARC logs and working copies of GIS coverages

common/sas Ungridded data directory

common/gis Ungridded GIS directory

Normally the define step should only be run once.

Generate Coverages

Coverages generates gridded coverages in the grid directories. It also generates ASCII gridded data files for each of the coverages selected. The input files used by the step are defined in the following environment variables:

COVDEF_IN Name of coverage definition file

The standard setting for COVDEF_IN is \$GISDB/covdef.in.

The remaining environment variables specify the coverages will be generated. The following list of environment variables was taken from the coverage names in the standard coverage definition table:

CENSUS	Extract census tract areas	(YES or NO)
AGRICULTURE	Extract agriculture areas	(YES or NO)
AIRPORTS	Extract airport counts	(YES or NO)

PORTS	Extract ports	(YES or NO)
RAILROADS	Extract railroad lengths	(YES or NO)
LAND_WATER	Extract land-water areas	(YES or NO)
URBAN_RURAL	Extract urban-rural areas	(YES or NO)
ROADS	Extract road lengths	(YES or NO)
FOREST	Extract forest areas	(YES or NO)

The coverages may be generated in one large step or in individual steps. The sample plan has three steps for generating coverages:

```
Generate all coverages except ROADS and CENSUS
Generate ROADS
Generate CENSUS
```

The rationale for this is that ROADS and CENSUS are large detailed coverages that may take a long time to generate.

The data files generated will be stored in the gridded data directory (\$EMS_HOME/gridspec/\$HPATHNAME/sas) The files generated for each coverage selected (except AIRPORTS and PORTS) will be:

```
<covname>.cy      County level grid data
<covname>.st      State level gridded data
```

AIRPORTS and PORTS coverages will only create the county level file, since they are POINT coverages. <covname> is a lower-case translation of the coverage name.

Compute Surrogates

Surrogates computes the spatial surrogates for the grid. The files used by the step are defined in the following environment variables:

```
FEATURE_SRG      Feature-surrogate table
CENSUS_DATA      Census data
FEAT_STCY_FRAC   Census feature state to county ratio table
SURROGATES       Spatial surrogate file
```

The spatial surrogate file is the output of the step. The other files are input. The standard settings for the input files are:

```

FEATURE_SRG      $GISDB/feature_srg.in
CENSUS_DATA      $GISDB/census_block_data
FEAT_STCY_FRAC   $GISDB/feature_stcy_ratios.dat

```

A sample setting for SURROGATES could be

```
$EMS_HOME/gridspec/${HGRIDNAME}/sas/ratios.dat
```

The remaining environment variables specify the coverages from which spatial surrogates will be computed. The coverages must have been previously generated in a coverages step. The following list of environment variables was taken from the coverage names in the standard coverage definition table:

CENSUS	Compute census surrogates	(YES or NO)
AGRICULTURE	Compute agriculture surrogates	(YES or NO)
AIRPORTS	Compute airport surrogates	(YES or NO)
PORTS	Compute port surrogates	(YES or NO)
RAILROADS	Compute railroad surrogates	(YES or NO)
LAND_WATER	Compute land-water surrogates	(YES or NO)
URBAN_RURAL	Compute urban-rural surrogates	(YES or NO)
ROADS	Compute road surrogates	(YES or NO)
FOREST	Compute forest surrogates	(YES or NO)

The spatial surrogate file must be generated in a single step with all coverages needed selected.

Test Surrogate File

```

# SMOKETOOL Spatial Surrogate Test Script.
# Parameters:
#   (1) Grid name

```

```

setenv EMS_HOME /m3aux/emis/susick/smoketool
setenv HGRIDNAME ${1}
setenv G_GRIDPATH /home/susick/test/GRIDS/m3grid_${HGRIDNAME}
setenv STUDY_PATH /home/susick/test
setenv GISDB $M3DATA/nostudies/gisdb
setenv GRID_DIR $EMS_HOME/gridspec/${HGRIDNAME}/sas

```

```

$M3MSTOOL/bin/smoktool_grid define
mv $STUDY_PATH/program.log $GRID_DIR/define.log
mv $STUDY_PATH/program.lst $GRID_DIR/define.lst

```

```
setenv COVER_DEF    $GISDB/covdef.in

setenv FEATURE_SRG  $GISDB/feature_srg.in
setenv CENSUS_DATA  $GISDB/census_block_data
setenv FEAT_STCY_FRAC $GISDB/feature_stcy_ratios.dat
setenv SURROGATES   $GRID_DIR/ratios.dat
```

```
setenv AGRICULTURE  yes
setenv AIRPORTS     yes
setenv CENSUS        yes
setenv FOREST        yes
setenv LAND_WATER    yes
setenv PORTS         yes
setenv RAILROADS     yes
setenv ROADS         yes
setenv URBAN_RURAL   yes
```

```
$M3MSTOOL/bin/smoktool_grid coverages
mv $STUDY_PATH/program.log $GRID_DIR/coverages.log
mv $STUDY_PATH/program.lst $GRID_DIR/coverages.lst
```

```
$M3MSTOOL/bin/smoktool_grid surrogates
mv $STUDY_PATH/program.log $GRID_DIR/surrogates.log
mv $STUDY_PATH/program.lst $GRID_DIR/surrogates.lst
```

Example Grid (m3grid_c_cmaqmet_36eval)

```
GDTYP_GD 0 LAMGRD3
P_ALP_GD 0 30.000000
P_BET_GD 0 60.000000
P_GAM_GD 0 -90.000000
P_XCENT_GD 0 -90.000000
P_YCENT_GD 0 40.000000
VGTPY_GD 0 VGSGPN3
VGTOP_GD 0 10000.0
VGTPUN_GD 0 Pa
VGLVUN_GD 0 none
VGLVS_GD 1 16
1.0 0.995 0.985 0.970 0.950 0.920 0.870 0.800 0.700 0.600 0.500 0.400 0.300 0.200 0.100 0.00
X3FACE_GD 1 16
0 0.005 0.015 0.03 0.05 0.08 0.13 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
```

GDNAME_GD 0 c_cmaqmet_36eval
GDDESC_GD 0 c_cmaqmet_36eval
XORIG_GD 0 -9.72000E+05
YORIG_GD 0 -1.620000E+06
XCELL_GD 0 3.600000E+04
YCELL_GD 0 3.600000E+04
GL_NROWS 0 81
GL_NCOLS 0 87
NPCOL 0 1
NPROW 0 1
NROWS 0 81
NCOLS 0 87
NTHIK 0 1
MNDIS 0 2
MEDIS 0 2
MSDIS 0 2
MWDIS 0 2
NLAYS 0 15